TRI-COUNTY PILOT STUDY

Job Order 75-315

Prepared By

Lockheed Electronics Company, Inc.

Aerospace Systems Division

Houston, Texas

Contract NAS 9-12200

For

EARTH OBSERVATIONS DIVISION

Science and Applications Directorate

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

July 1976

LEC-8657

TRI-COUNTY PILOT STUDY

Job Order 75-315

PREPARED BY

C. A. Reeves, T. W. Austin, A. G. Kerber

APPROVED BY

LEC

J. E. Davis, Manager
Earth Observations Exploratory Studies Department

NASA/JSC

R. E. Joosten, Deputy Manager
Forestry Applications Program

USDA, Forest Service

F. P. Weber, Manager
Forestry Applications Program

Prepared By

Lockheed Electronics Company, Inc.

For

Earth Observations Division

Science and Applications Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

July 1976

LEC-8657

iii

PREFACE

The Tri-County Pilot Study was initiated to develop and test procedures for computer-aided analysis of satellite data for performing area inventories of forestland and rangeland features.

Assistance in the preparation of this report was given by Dr. R. W. Douglass, Mr. I. E. Duggan, and Mr. W. H. Parkhurst. Special assistance was provided by Dr. E. P. Kan, whose critical review and suggested improvements were invaluable to the completion of the Tri-County Pilot Study.

iv

PAGE INTENTIONALLY BLANK

iv

PRECEDING PAGE BLANK NOT FILMED.

CONTENTS

Vi

Vi

TABLES

FIGURES

1. SUMMARY

The Tri-County Pilot Study (TRICPS) of the Forestry Applications
Project (FAP) investigated the feasibility of using computer-
aided classification of Landsat-1 [formerly Earth Resources Tech-
nology Satellite (ERTS)] data, with supporting aircraft imagery
and ground survey data, for performing forestland and rangeland
inventories. The planning and execution of the study required
three project personnel and was completed in 1 year.

An area inventory by administrative boundaries was performed for
three southeast Texas counties (Montgomery, Walker, and San
Jacinto) totaling 0.65 million hectares (1.6 million acres). Data
from Landsat-1 taken in November 1972 and May 1973 were used for
the inventory. Two approaches were found to be feasible, one a
wall-to-wall approach where every picture element (pixel) of the
data was classified and the other a sampling approach where a
subset of the data was classified and used to estimate the total
area. The General Electric Interactive Multispectral Image
Analysis System (Image 100) was used for computer classification.

The inventory was performed using a two-level hierarchy. Level I
was divided into forestland, rangeland, and other land (nonforest-
land and nonrangeland). Forestland was separated into Level II
categories: pine, hardwood, and mixed pine/hardwood; rangeland
was not separated further.

The results of the study consisted of area statistics for each
county and for the entire study site for pine, hardwood, mixed
pine/hardwood, rangeland, and other land. In addition, color-
coded county classification maps were produced for the May data
set, and procedures that were developed and tested were documented.

The area inventory results were evaluated using three different
methods: one was a comparison with statistics from the Forest

Service (FS) and Soil Conservation Service (SCS) of the U.S. Department of Agriculture (USDA); another was a test field evaluation of 1 percent of Walker County, comparing color infrared aircraft photography with computer classification maps; and the third was a random sample evaluation of the entire study site comparing aircraft photography and computer classification maps, from which the confidence interval and probability of correct classification were calculated.

The accuracy of the results varied considerably depending on the evaluation method. The accuracy figures for the evaluation with historical data varied among the several classes and the two data sets. The areal measurements from the May data set showed 5.8-, 6.1-, and 11.1-percent differences, respectively, for pine, hardwood, and range, relative to FS and SCS statistics. Classification accuracy figures were highest for the test field evaluation method, due in part to its simplicity of execution and the absence of boundary pixels, which tend to decrease accuracy. Figures from the Walker County May data set showed 95-, 76-, and 71-percent accuracies for pine, hardwood, and range, respectively. The random sample evaluation produced accuracies much lower than those from the test field evaluation. These lower accuracies were due in part to registration/identification problems and were considered to be unreliable. It was concluded that the mapping of mixed pine/hardwood stands from Landsat data was not feasible using the procedures of TRICPS. The study site did not provide enough homogeneous mixed stands to produce a meaningful signature.

2. INTRODUCTION

The TRICPS was initiated to study the feasibility of computer-aided analysis of satellite data for performing forestland and rangeland inventories. Three southeast Texas counties (Montgomery, Walker, and San Jacinto) were selected as a study site to develop and test procedures for (1) an acreage inventory of features by administrative boundaries and (2) a sampling analysis approach. These procedures and any modifications would be used in the design and implementation of future state, regional, and national inventories.

2.1 BACKGROUND

The FAP was initiated in 1971 at the Lyndon B. Johnson Space Center (JSC) of the National Aeronautics and Space Administration (NASA), with the objective of supporting the Southern Region of the FS in the development and testing of remote sensing analysis techniques to perform forest resource inventories (ref. 1). Satellite and aircraft data analysis has been demonstrated to hold promise in effective resource management. Investigators such as Heller (ref. 2) used autumn Landsat data and computer classification techniques to separate pine and hardwood with accuracies ranging from 42 to 81 percent. Erb (ref. 3) reported separating summer Landsat data into pine and hardwood with 91- and 70-percent accuracy, respectively.

The need for remote sensing analysis techniques to monitor large area forestland and rangeland resources was intensified by the passage of Public Law 93-378 in August 1974, the "Forest and Rangeland Renewable Resources Planning Act of 1974." This law authorized the Secretary of Agriculture to prepare and update periodically a Renewable Resources Assessment and Program for the United States, its territories, and possessions. The first such assessment and program were published in December 1975 (refs. 4 and 5). The next assessment and program will be due in 1979 and 1980, respectively.

In response to the need for the development of remote sensing analysis techniques, TRICPS was initiated by FAP to develop and test procedures to inventory forestland and rangeland resources in preparation for future state, regional, and national inventories.

One highlight of TRICPS was to test the feasibility of a forestland and rangeland inventory by administrative boundaries. Such attempts were not operationally feasible until the advent of state-of-the-art computer and optics hardware, which allow ancillary data in the form of maps to be input to the computer. The Image 100 purchased and installed at JSC in 1974 is such a system, and it was made available for the study.

Another highlight of the study was to develop and test a sampling analysis approach to inventory large areas. Although the three-county study area of 0.65 million hectares (1.6 million acres) was not large enough to warrant a cost-effective sampling inventory, such an attempt will prepare FAP for future state, regional and national inventories where a sampling strategy is unavoidable.

2.2 OBJECTIVES

TRICPS had the following specific objectives:

- To develop and test computer-aided classification procedures to inventory forestland and rangeland areas by administrative boundaries

- To develop and test procedures for using a sampling analysis approach to inventory forestland and rangeland areas

The features of the inventory are discussed in section 2.4.

2.3 SCOPE

The TRICPS area, composed of Montgomery, Walker, and San Jacinto Counties, covers 0.65 million hectares (1.6 million acres),

approximately 65 kilometers by 90 kilometers (40 miles by 55 miles). (Refer to section 3 for additional description of the study site.)

Satellite data from Landsat-1 were used for analysis. Two sets of data, representing the spring and autumn seasons, were used. In addition, supporting aircraft photography and ground survey data were used for site familiarization, training field selection, and evaluation of results.

The project planning, execution, and analysis were completed in the calendar year of 1975 and required three project personnel plus supportive cartographic and photographic services.

Analysis equipment used in the project consisted principally of the Image 100 with support from the Passive Microwave Imaging System (PMIS) Data Analysis Station (DAS), the computer terminal connected to the Laboratory for Applications of Remote Sensing (LARS) of Purdue University, the Kargl reflecting projector, and the Zoom transferscope.

A two-level hierarchy was adopted from the Anderson-Hardy-Roach land use classification system (ref. 6). Level I was comprised of forestland, rangeland, and other land. Only forestland was further analyzed to Level II — pine, hardwood, and mixed pine/hardwood. (Refer to section 2.4.)

Two analysis approaches were used, one a wall-to-wall approach in which every pixel of the data was analyzed and the other a sampling approach in which only a subset of the data was studied.

The results of the study included areal statistics of forestland and rangeland features per county, county classification maps, classification map accuracies with their evaluation statistics, and documentation of analysis procedures used in this study.

2.4 ANALYSIS LEVELS

In the TRICPS study only forestland and rangeland features were classified; the remaining unclassified features were considered "other" land. Specifically, the inventory was performed using a two-level hierarchy (table 2-1):

Level I - separation of forestland, rangeland, and other land (nonforestland, nonrangeland)

Level II - separation of forestland into pine, hardwood, and mixed pine/hardwood

Initially, rangeland was to be separated into Level II categories: improved range and nonimproved range. However, it became apparent early in the study that it was difficult to discriminate between the two categories of range on aircraft data. Since aircraft data were used to pick training fields, project personnel felt that improved range training fields could not be discriminated from non-improved range training fields. In addition, no ground-truth maps of vegetation cover were available for the rangeland in the study area. Therefore, it was decided to separate the forestland only and not to further analyze rangeland into Level II categories.

Standard definitions of features of interest accepted by the Society of American Foresters (refs. 7 and 8), the FS survey (ref. 9), and the U.S. Department of the Interior (USDI) Geological Survey (GS, ref. 6) were considered incompatible with remote sensing applications, and therefore the definitions were modified as follows:

- Forestland - land of 0.4 hectare (1 acre) and larger in size supporting a stand of trees whose crowns cover more than 10 percent of the area

- Rangeland - land, excluding forestland, that produces forage for animal grazing

TABLE 2-1.— HIERARCHY OF FEATURES STUDIED
IN THE TRI-COUNTY PILOT STUDY

LEVEL I	LEVEL II
Forestland[1]	Pine Hardwood Mixed pine/hardwood
Rangeland Other land	

[1]In the TRICPS study, only forestland and rangeland features were classified; the remaining unclassified land was "other" land.

- Other land – nonforestland and nonrangeland, implying cropland, urban areas, water, and miscellaneous

- Pine – trees of the genus *Pinus*, having evergreen and needle foliage; a stand comprising more than 50-percent pine in the overstory

- Hardwood – trees of the subdivision Angiospermae, generally having broadleaved and deciduous foliage; a stand comprising less than 25-percent pine in the overstory

- Mixed pine/hardwood – a stand comprising 25- to 50-percent pine in the overstory

3. STUDY SITE

The study area (figure 3-1) is part of the "east Texas piney woods," also called "flatwoods"; it lies in the Gulf Coastal Plains physiographic province. Topography is flat to gently rolling, with a sandy topsoil over a heavy clay subsoil with clay outcrops.

Forest vegetation generally consists of shortleaf pine (*Pinus echinata* Mill.) on ridges and upper slopes, and loblolly pine (*Pinus taeda* L.) and hardwoods on the lower slopes and in the bottoms. The hardwoods are primarily laurel oak (*Quercus laurifolia* Michx.), willow oak (*Quercus styraciflua* L.) and nuttal oak (*Quercus nuttallii* Palmer). On some high, dry sites post oak (*Quercus veluntina* Lam.) predominates. Further descriptions of these timber types can be found in reference 9.

The climate is subtropical in nature with hot, humid summers and mild winters. The average annual rainfall ranges from 107 to 127 centimeters (42 to 50 inches).

The study site contains the Sam Houston National Forest, the Jones State Forest, private forests, and rangeland. The area also includes two large lakes (Lake Conroe and portions of Lake Livingston), and several small rural communities. The Sam Houston National Forest comprises approximately 65,000 hectares (160,000 acres). It is divided into two ranger districts, the Raven District [40,000 hectares (100,000 acres)], and the San Jacinto District [25,000 hectares (60,000 acres)]. The Jones State Forest contains approximately 700 hectares (1700 acres), whereas private forests cover 400,000 hectares (1,000,000 acres). Approximately 120,000 hectares (280,000 acres) are rangeland.
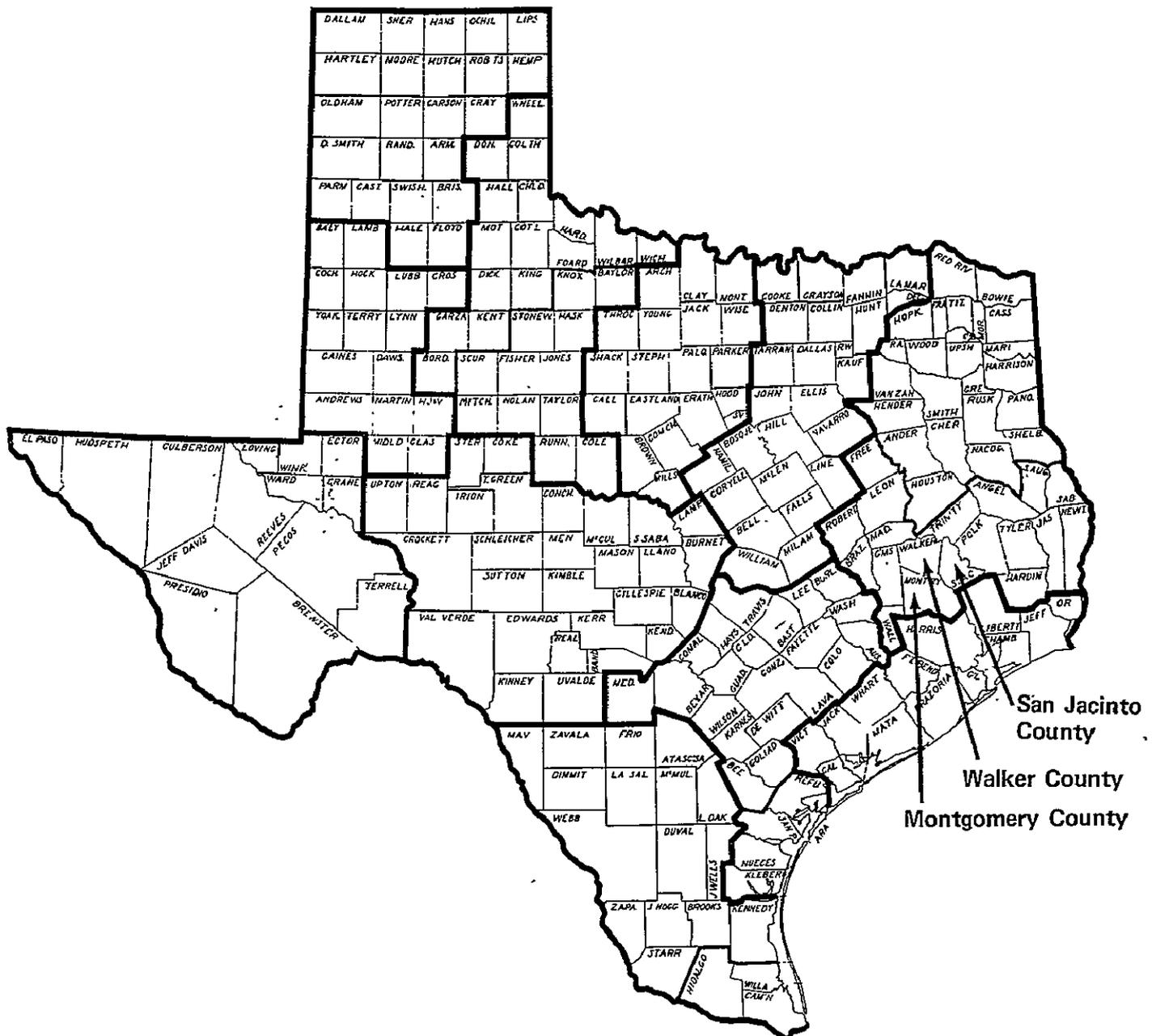
Figure 3-1.— Texas county map showing the Tri-County Pilot Study area of Walker, Montgomery, and San Jacinto Counties.

4. DATA UTILIZATION AND OUTPUT PRODUCTS

4.1 LANDSAT-1 DATA

TRICPS processed and classified Landsat-1 multispectral scanner (MSS) data to obtain areal estimates and classification maps. The Landsat-1 MSS has four channels, or bands, which cover the following spectral regions: band 4 (500 to 600 nanometers), band 5 (605 to 700 nanometers), band 6 (693 to 800 nanometers), and band 7 (810 to 1100 nanometers). These bands are also known as channels 1, 2, 3, and 4. Details on the characteristics of Landsat sensor systems and platform can be found in reference 10.

Two sets of Landsat-1 data were individually computer classified; they were taken in November 1972 (I.D. 1217-16254) and May 1973 (I.D. 1289-16254). The May data set was chosen to represent a spring date with a leaf-on forest condition. This season has been shown to be one of the best for discrimination between pine and hardwood in the southeastern United States (ref. 11). The November data set was chosen as a comparison date representing a dormant forest condition. In addition, rangeland identified in the May data set was in a green, spring-growth condition, whereas rangeland identified in the November data set was in a brown, dormant-vegetation condition.

4.2 AIRCRAFT IMAGERY

Color infrared photography was used to aid in the selection of training fields and evaluation of classification results. Photography obtained by NASA aircraft (Mission 197 in April 1972 and Mission 239 in June 1973, at respective scales of 1:120,000 and 1:60,000) was utilized. These photographs were taken using Kodak film type 2443 with a Wratten 12 filter, which covers an effective spectrum of approximately 510 to 900 nanometers.

4.3 ANCILLARY DATA

The GS topographic quad sheet (1:250,000, NH 15-4, 1963) for
Beaumont, Texas, provided the base to which Landsat data were
registered. This registration permitted input of administrative
boundaries using the Image 100.

Compartment maps provided by the FS for the two ranger districts
of the Sam Houston National Forest were used as aids in locating
pine, hardwood, and mixed pine/hardwood stands for the selection
of training fields. These maps show the identification and loca-
tions of stands within each compartment.

The Southern Forest Experiment Station in New Orleans, Louisiana,
provided FAP with 1975 areal statistics for forest types on a
per-county basis (ref. 12). Areal statistics for rangeland for
1967 were obtained from the SCS (ref. 13). These statistics
provided the base by which the TRICPS statistics were compared
and evaluated. Rainfall data for the period surrounding the
May 1973 Landsat overpass were obtained from the National
Oceanic and Atmospheric Administration (NOAA) weather station
located at Huntsville, Texas (ref. 14).

4.4 GROUND DATA

Ground checks were made in the summer of 1975 by FAP personnel
with the help of the FS to familiarize the photointerpreters with
the general appearance of selected features within the study site.

4.5 OUTPUT PRODUCTS

Output products for the wall-to-wall classification of both data
sets included (1) areal statistics by class for each county, (2)
county classification maps, (3) map accuracies with their evalua-
tion statistics, and (4) documentation of data analysis procedures
that were developed and tested. The output product for the samp-
ling approach consisted of areal estimates for each class within
the entire study site.

From the map evaluation process, the measures of probability of correct classification and confidence interval were calculated. One-class classification maps from the line printer were produced for both data sets and were used in the evaluation of results. Composite, color-coded, county classification maps made from the printouts were produced for the May data set. Refer to section 6 (figs. 6-1, 6-2, and 6-3) for examples of the classification maps. Due to limitations of time and manpower, no composite classification maps were produced for the November data set.

5. TECHNICAL APPROACH

5.1 OVERVIEW

The TRICPS was a computer-aided inventory of forest and range features using Landsat-1 MSS data. One set of spring data and one set of autumn data (section 4.1) were analyzed individually using the Image 100 to separate the features of interest. Each of these data sets was registered to a topographic map base so that administrative boundaries could be input, and thus the basis for inventory could be established.

Two approaches were used to perform the inventory in TRICPS. The first was a wall-to-wall approach in which every pixel of Landsat data was classified, producing classification maps and areal estimates for the features of interest. The second, a sampling approach, produced areal estimates from the classification of a subset of the study area data.

The classification in both approaches was performed using signatures produced by training fields. The training fields were selected, their signatures extracted, and the classification executed on the Image 100.

5.2 DATA ANALYSIS STEPS

All of the analysis steps in TRICPS had to be performed within the data handling constraints of the Image 100. For processing, the Image 100 is limited essentially to 507 lines and pixels and 493 lines for display on the screen instead of the 512- by 512-pixel/ line design of the video screen (ref. 15).

Therefore, the Landsat data were divided into segments containing 500 lines and 500* pixels. Fourteen of these segments were required to cover the three counties.

---

*The constraint of 493 lines on the Image 100 display screen was not realized at the time of TRICPS analysis, hence the choice of 500- by 500-pixel/line segments.

The basic procedures planned for the wall-to-wall processing approach are presented in this section. (Refer to appendixes B through G for detailed steps of these procedures.) The sampling approach differed only in that classification was performed on specified portions of the study area, so none of the procedural steps prior to classification (registration, boundary inputs, and signature acquisition) needed to be repeated for the sampling approach. Figure 5-1 is a flow diagram depicting the general processing approach and showing processing steps performed on the Image 100, along with output products.

5.2.1 REGISTRATION PREPARATION

To prepare for the registration of the Landsat data to a map base, control points were located on both the Landsat data and the map. The exact locations for points in Landsat line and pixel numbers and geographic (or arbitrary x,y) coordinates were input to a least-squares-fit program. This program was run iteratively and control points with large deviations rechecked or deleted until the overall fit was acceptable. Coefficients from the least-squares-fit program and a scale factor were input to a program to calculate the rotation factor to deskew the Landsat data. In addition, the positions of each of the 14 segments, 500 lines by 500 pixels on the Landsat data, were determined.

5.2.2 REGISTRATION

Registration of the Landsat data to the topographic base map was performed by deskewing the data using the rotation factor calculated during registration preparation and by properly scaling (downward) the pixels in the line of flight (i.e., across scan lines). Each segment of data was input based on starting and ending line and pixel numbers calculated during registration preparation and deskewed based on the rotation factor. The

ACQUISITION OF
LANDSAT-1 MSS DATA
NOVEMBER 1972 AND
MAY 1973

FILMING OF REGISTERED
LANDSAT DATA ON
PMIS DAS

CLASSIFICATION

EVALUATION OF
CLASSIFICATION
RESULTS

REGISTRATION OF THE
LANDSAT DATA TO A
TOPOGRAPHIC BASE MAP

FILM TRANSPARENCIES

CLASSIFICATION GRAY MAPS

ACCURACY ESTIMATES

REGISTERED TAPES

SELECTION OF
TRAINING FIELDS

COLOR COMPOSITE MAP

PROBABILITY OF
CORRECT CLASSIFICATION

INPUT OF ADMINISTRATIVE
BOUNDARIES USING THE
REGISTERED TOPOGRAPHIC MAP

REGISTERED
TAPE WITH
COUNTY
BOUNDARY

CLASS
SIGNATURE
ACQUISITION

OUTPUT PRODUCTS

PROCESSING PERFORMED ON GE100

Figure 5-1.— Flow diagram for the general processing approach in TRICPS.

downward scaling was achieved by inputting less than 500; in fact, 369 Landsat lines were displayed as 500 Image 100 lines.

With the Landsat data registered to the map, a portion of the map corresponding to each segment was input to the Image 100. The proper scale and rotation adjustments were made so that a topographic map transparency could be input in place of band 4 of the Landsat data using the input scanner. This was achieved by moving the scanner camera to the proper height and by rotating the light table on which the transparency was mounted. Channel 1 of the Image 100 then displayed the map for the input of county (administrative) boundaries.

5.2.3   INPUTTING BOUNDARIES

The topographic map (1:250,000 scale) was input in place of band 4 of the Landsat data during registration. By displaying the map, administrative boundaries were traced and stored so that areas outside the counties could be deleted later from the classification. This would allow classification maps and area figures to be produced on a per-county basis.

5.2.4   SIGNATURE ACQUISITION

Prior to selection of training fields, field trips were taken to familiarize photointerpreters with the appearance of features on aerial photography. Using this familiarization and FS stand maps, areas of pine, hardwood, mixed pine/hardwood, and range were delineated on photographs and transparencies of the Landsat data.

From these locations, training fields for the four classes were selected on the Image 100 screen using the May data. Five, ten, and fifteen training fields per class for each county were used.

One-dimensional* signatures were extracted from training fields for each class in each county. For each signature, 1 percent of the pixels was truncated from the upper and lower spectral bounds for each channel. The signatures for the four classes were expected to have overlapping spectral bounds in some cases. In order to avoid pixels being classified as more than one class, overlap had to be resolved in at least one channel for each pair of class signatures. For the channels where overlap had to be resolved, the spectral values in which overlap occurred were assigned to the class that contained more pixels.

Four class signatures were developed for each county, as well as a composite of all three counties, using the May data set. As nearly as they could be located, the same training fields were used for the November data to produce county and composite signatures for all classes.

5.2.5 CLASSIFICATION

The one-dimensional signatures for the four classes, pine, hardwood, mixed pine/hardwood, and range, were input to the Image 100 to produce a classification of the Landsat data. After the classification was complete, the unclassified pixels were assigned to the class called "other." Areas outside the tri-county site were subtracted from the classification so that county maps and statistics could be produced.

---

*A "one-dimensional" signature is extracted on the Image 100 by the recording of the highest and lowest spectral values or "bounds" in each channel for all the pixels in the training area. In the subsequent classification, the Image 100 checks the spectral values of each pixel to be classified and, if the values fall within the spectral bounds of the one-dimensional signature, then that pixel is assigned to the signature class.

Area statistics per class were calculated from pixels per class using the following formula:

$$A = P \times \frac{L}{500} \times \frac{S}{500} \times 0.45* \text{ hectare}$$

where

A = area in hectares per class

P = number of pixels per class

L = number of lines of Landsat data loaded on 500 lines of the Image 100, L = 352 for this study

S = number of samples (points) of Landsat data loaded on 500 points of Image 100, S = 500 for this study.

5.2.6 EVALUATION

Three evaluation methods were planned to check classification results. One method was to compare areal estimates for each class obtained from classification with FS survey statistics for each county. The other two methods were to evaluate the accuracy of classification maps; they involved locating sample plots on aerial photography and classification maps and comparing the two to obtain accuracy figures. One evaluation was based on the selection of test fields by a photointerpreter and the other on randomly located plots.

These three methods were used to evaluate the wall-to-wall analysis results. For the sampling approach, only the first evaluation method by areal estimates was used because the approach produced areal statistics only; it did not produce maps.

---

*0.45 hectare is the area of one Landsat pixel, based on a pixel size of 57 by 79 meters.

5.3 DISCUSSION OF WALL-TO-WALL APPROACH

This section discusses the rationale for some of the decisions made in the actual TRICPS analysis. Specifically, the registration, signature acquisition, classification, and evaluation steps of the approach required decisions worthy of discussion.

5.3.1 REGISTRATION

5.3.1.1 Geometric Correction of the Landsat Data

In order to superimpose administrative boundaries on Landsat data using the Image 100, the data must be reformatted to a correct geometric representation. This problem arises because Landsat pixels are nonsquare but are represented as square on the Image 100 video display. For proper registration, a square area on the video screen must represent a square ground area. This can be accomplished by periodically repeating Landsat lines or deleting pixels. Because deleting pixels eliminates useful data, the addition of lines was the procedure chosen for use in TRICPS. (Refer to appendix B for details on calculation of the proper number of lines to input.)

5.3.1.2 Deletion of Landsat Data for Topographic Map Input

The input of a map base to the Image 100 requires the deletion of one data channel. In TRICPS, band 4 was chosen to be deleted because (1) the undesirable atmospheric effects are most severe in this channel and (2) the absolute signal levels are lower in this band than in either of the infrared channels. Furthermore, examination of the May data set showed high noise levels in this channel.

5.3.2 SIGNATURE ACQUISITION AND CLASSIFICATION

5.3.2.1 Gain and Radiometric Resolution

During input, the Landsat data were scaled upward with gain factors of 2, 2, 2, and 4 for bands 4, 5, 6, and 7, respectively,

in order to fill the entire range of 256 gray levels in the Image 100 memory. For efficient processing, the user must decide the proper radiometric resolution or number of discrete levels into which these 256 gray levels are to be divided. The choices for resolution are powers of 2, with 64 (2^6) the default value if none is specified. If too few levels are used, the usefulness of the data in separating different features is diminished. However, a resolution is usually reached, less than 256, above which essentially no better separation between feature signatures can be obtained (ref. 15). Using a resolution above this will also have the adverse effect of requiring additional processing time. Iterative analyses of signatures at various resolutions were made to decide on the optimal resolution.

In TRICPS, stands of pine and hardwood were identified from photography for a portion of the study area. These delineated stands were used to check the accuracy of maps produced from classifications using different resolutions. At a resolution of 32 (the 256 gray levels assigned to 32 resolvable levels, or 8 gray levels for each resolvable level), pine and hardwood could not be differentiated, resulting in gross errors in classification. At a resolution of 64, classification error was about 1 percent for the two classes. Higher resolutions did not improve classification accuracy; therefore, 64 was decided upon as the appropriate radiometric resolution. (Refer to appendix A for details of this procedure.)

5.3.2.2 Number and Size of Training Fields

The use of 5, 10, and 15 training fields for each class in each county was planned to investigate the effects of variations in number and distribution of training fields on classification accuracy. The maximum of 15 training fields per class for each county was based upon practical considerations of ground checks. It was estimated that about 10 training field locations could be

checked per day, provided the locations were near transportation networks. From this estimate, all training fields could be ground checked in about 3 weeks, which was considered a reasonable amount of field checking for TRICPS.

The size of the training fields was chosen to be 36 pixels (about 12 hectares), unless this size was too large to fit in a potential training area. Hardwood stands often occurred as sinuous areas along rivers and streams, and range areas were often interspersed with clumps of trees or small water areas. Therefore, it was necessary to sometimes reduce the training field size to fit the particular area.

5.3.2.3 Truncation of the Signatures

Each signature was truncated to remove 1 percent of the training area pixels from both the upper and lower spectral limits for each channel. The decision to eliminate these pixels was arbitrary but was based on the belief that such procedures help eliminate any pixels included in a training area which in fact were not of that class.

5.3.3 EVALUATION PROCEDURES

5.3.3.1 Test Field Evaluation

As an initial check of the locational accuracy of the classification, a test field approach was developed. Aerial photography over one county was used to locate potential test fields containing only one class each. From these potential test fields, a sample was selected equaling 1 percent of the county by area. The number of test fields for each class in the sample was proportional to the area of each class in the county as determined from classification statistics.

The identities of the test fields were known from previous interpretation of the photographs. They were then located on classification maps to determine their classifications. Calculation of percent accuracy for each class was made by comparing the two.

5.3.3.2 Random Sample Evaluation

In order to calculate the accuracy with a confidence interval
(hence, the reliability) of the classification map accuracy,
another evaluation method was used, which involved checking the
classification of randomly selected plots. The computer clas-
sification map was evaluated against interpreted color infrared
photography. Areas not classified as one of the four classes
were called "other" and included as another class in the eval-
uation, because the classification map is actually a five-class
map containing pine, hardwood, mixed pine/hardwood, range, and
"other."

The randomly selected plots were located on the classification
maps and the maps registered to the photographs for identifica-
tion of each plot. The comparisons of computer classification
to photographic identification were tabulated and two-class map
accuracies, $P(C_i, \overline{C}_i)$, such as pine (C_i), nonpine $(\overline{C}_i)$, calculated
using the following equation (ref. 16).

$$P(C_i, \overline{C}_i) = \frac{C_i + \overline{C}_i}{N}$$

where

C_i = plots correctly classified as class i

$\overline{C}_i$ = plots not class i correctly classified

N = total number of samples

The $P(C_i, \overline{C}_i)$ for each class would then be used to calculate the
M-class accuracy (ref. 16):

$$P(C_1, C_2, \ldots, C_M) = 1/2 \, [P(C_i, \overline{C}_i) + P(C_2, \overline{C}_2)$$
$$+ P(C_M, \overline{C}_M) - (M - 2)]$$

where

M = number of classes

= 5 in TRICPS output maps

Because of the randomness (independence) of the sample plots, the map accuracy estimates approximate a binomial distribution when N is large. This allows the calculation of a confidence interval for the map accuracy.

The 90-percent confidence interval for the M-class map accuracy is expressed by:

$$\left[P(C_1,C_2,\ldots,C_M) - \Delta, \; P(C_1,C_2,\ldots,C_M) + \Delta\right]$$

$$\text{where} \quad \Delta = 1.64\sqrt{\frac{P(C_1,C_2,\ldots,C_M)\,[1 - P(C_1,C_2,\ldots,C_M)]}{N}}$$

and 1.64 is the "t-value" for the 90-percent confidence interval (ref. 17).

For a distinction between the map accuracy $P(C_1,C_2,\ldots,C_M)$ and the commonly used class accuracy (such as that used in the "test field evaluation" discussed in section 5.3.3.1), see reference 16.

5.4   DISCUSSION OF THE SAMPLING APPROACH

The sampling approach developed in TRICPS was used to investigate the accuracy of areal estimates produced by sampling portions of the study area. These portions, called primary sampling units (PSU's), became intensive study sites for wall-to-wall classification, which was then used to estimate areas for each class for the entire test site. The classification of the PSU's was based on signatures developed in the wall-to-wall approach.

At the outset of TRICPS, the sampling procedure was required to produce estimates of the wall-to-wall area figures at a 90-percent confidence level with an allowable error of ±5 percent for all feature classes. In addition, the size of a PSU was chosen to be 35 by 35 pixels. An optimal PSU size is actually governed by the variability between PSU's and relates to the efficiency and complexity of the design of the sampling procedure. Because this relationship was not available for TRICPS, an arbitrary size was chosen.

The number of PSU's required to produce areal estimates of a feature class at a given confidence level and allowable error can be determined if the sample is random. The following formula defines this relationship:

$$n = \frac{t^2 s_B^2}{E^2} .$$

where

n   = number of PSU's

S_B^2 = between-PSU variance for a class

E   = confidence interval half range, i.e., allowable error
       (0.05 prescribed for TRICPS)

t   = value of the t-distribution to achieve the specified confidence level (1.64 prescribed for TRICPS at 90-percent confidence level)

Since the between-PSU variance S_B^2 varies from class to class, this formula must be used independently to determine the required number of PSU's for each of the five classes (four classes plus "other" class). Then, the largest of these n's was used as the final selection of the number of PSU's to be used in the TRICPS sampling analysis. However, theoretical analysis of the

formula suggested that the n corresponding to the majority class
(pine for the TRICPS study area) would likely be largest.*

The between-PSU variance for each feature class could be estimated
by performing a preliminary analysis where K random PSU's were
classified. The proportions of each class in the $j th$ PSU, p_j,
could be calculated by counting pixels. S_B^2 could then be esti-
mated for each class as

$$S_B^2 = \frac{1}{K-1} \sum_{j=1}^{K} (p_j - \bar{p})^2$$

where

$$\bar{p} = \frac{1}{K} \sum_{j=1}^{K} p_j$$

An alternate form for S_B^2 is $\frac{1}{K-1} \sum_{j=1}^{K} (p_j^2 - \bar{p}^2)$.

In the TRICPS analysis, K was chosen to be 20. The formulas for
S_B^2 and n were used to evaluate the sample sizes (i.e., number
of PSU's) required for each class, hence the maximum n used for
the TRICPS sampling approach.

After n was determined, this number of PSU's was randomly located
without replacement in the TRICPS site. Each was classified
entirely; i.e., every pixel, using signatures derived in the wall-
to-wall approach. Then the average proportion of each class with-
in all the PSU's was calculated. These proportions were then
expanded to the total study area to produce final areal estimates
for each class.

6. ANALYSIS RESULTS

6.1 WALL-TO-WALL APPROACH

The primary method used in this study for obtaining area inventory results was the wall-to-wall approach in which every pixel of the data was classified. The remainder of this section covers in detail the signature acquisition, classification results, and evaluation results of the wall-to-wall approach.

6.1.1 SIGNATURES

During the selection of training fields, it was discovered that the original plan for the number and size of the training fields could not be met.* In the case of rangeland, there were less than 15 sufficiently large areas in each county for reliably placing training fields on the Image 100 screen. Because of this, some of the fields had to be reduced from the 36-pixel size in order to fit in a known area.

The mixed training fields presented more problems. It was not possible in two of the three counties to locate enough good mixed areas that were sufficiently large to develop a meaningful signature. Most of the areas identified as mixed by the FS or photo-interpretation were actually conglomerates of smaller pine and hardwood stands rather than truly homogeneous mixed areas. Signatures extracted from such areas would include spectral values characteristic of both pine and hardwood. However, nine training fields, totaling 36 pixels, were identified in San Jacinto County Because of the limited reliability of such a small training population, this signature was extracted and used for classification in the May data only. The location of these small training fields on the November data would have been more difficult because the

---

*Time and manpower limitations prevented thorough investigation and documentation of the 5 and 10 training field cases for each class and county. A cursory examination of 5 training fields for pine and hardwood showed that they produced signatures very similar to those of 15 fields.

forest classes for this date were not visually as separable on the Image 100 screen as they were for the May date.

As a result, 48 pine, 45 hardwood, 23 range and 9 mixed pine/ hardwood training fields were selected for the entire study area for the May data. The signatures produced by these fields for the May and November data sets, including 1-percent truncations (section 5.3.2.3), are presented in tables 6-1 and 6-2. The county signatures for each class were very similar, indicating that any one of them could be used to classify another county and produce a similar classification.

As expected, the pine, hardwood, and mixed signatures did have overlapping spectral bounds in several cases. For pine and hardwood, which overlapped in all channels, the overlap had to be resolved in one channel to prevent dually classified pixels. Spectral limits for each signature in Landsat band 6 for the May data and band 5 for the November data were truncated, assigning the spectral values to the class having more pixels with that value. The mixed signature again provided a special problem because its spectral values fell entirely within pine or hardwood signatures in every channel. To allow for a mixed signature which would guarantee that mixed pixels would only be classified as mixed, pine and hardwood signatures were both truncated again in Landsat band 6. Decisions on truncation boundaries were arbitrary at the time of analysis.

The signatures resulting from these truncations were subsequently used for classification of the Landsat data. Table 6-3 lists these signatures.

TABLE 6-1.— ONE-DIMENSIONAL CLASS SIGNATURES FOR EACH COUNTY AND A COMPOSITE
FOR THE TRICPS STUDY SITE, BASED ON THE MAY LANDSAT DATA

Class	County (or composite)	Number of training fields	Number of pixels	Landsat channel		
				5	6	7
Pine	Walker	16	576	6-8	16-23	18-28
	Montgomery	16	576	6-8	14-22	15-26
	San Jacinto	16	574	6-8	16-23	19-27
	Composite	48	1726	6-8	15-23	18-27
Hardwood	Walker	15	396	6-8	22-30	28-36
	Montgomery	15	508	6-8	21-30	26-37
	San Jacinto	15	452	5-7	23-30	28-37
	Composite	45	1356	6-8	22-30	27-37
Range	Walker	8	144	9-17	22-30	24-35
	Montgomery	7	252	10-13	21-32	22-38
	San Jacinto	8	176	10-16	20-31	22-37
	Composite	23	572	9-14	21-32	23-37
Mixed	San Jacinto	9	36	6-8	22-24	26-29

TABLE 6-2.— ONE-DIMENSIONAL CLASS SIGNATURES FOR EACH COUNTY AND A COMPOSITE FOR THE TRICPS STUDY SITE, BASED ON THE NOVEMBER LANDSAT DATA

Class	County (or composite)	Number of training fields	Number of pixels	Landsat channel		
				5	6	7
Pine	Walker	16	576	4-6	8-12	9-13
	Montgomery	16	576	4-6	9-12	10-13
	San Jacinto	15	540	4-6	8-11	9-13
	Composite	47	1692	4-6	8-12	9-13
Hardwood	Walker	15	396	5-7	8-13	10-15
	Montgomery	15	540	5-7	9-12	9-14
	San Jacinto	15	452	5-7	9-12	9-15
	Composite	45	1388	5-7	9-12	9-14
Range	Walker	8	120	8-16	11-17	11-18
	Montgomery	7	252	8-13	12-18	13-21
	San Jacinto	8	176	8-13	11-20	11-21
	Composite	23	548	8-13	11-18	12-21

TABLE 6-3.— SPECTRAL SIGNATURES USED FOR CLASSIFICATION
OF MAY AND NOVEMBER DATA SETS

Class	May data set			Class	November data set		
	5	6	7		5	6	7
Pine	6-8	16-23	18-27	Pine	4-5	8-12	9-13
Hardwood	6-8	24-30	27-38	Hardwood	6-7	9-12	9-13
Mixed	6-6	23-24	26-28	Mixed	-	-	-
Range	10-13	23-30	25-35	Range	8-13	11-18	12-21

6.1.2 CLASSIFICATION RESULTS

The following results were obtained from the classification procedures.

6.1.2.1 Area Statistics

Based on classification in the wall-to-wall approach, the areas for each class and the total areas are listed by county for the May and November data sets in tables 6-4 and 6-5. Aggregated areas by class for the entire site for both data sets are presented in table 6-6.

The ability to inventory by administrative (specifically, county) boundaries has been demonstrated. The estimates of "total area in county," produced by the input of county boundaries and subsequent subtraction of areas outside counties, varied less than 2 percent on the average for both the May and November dates. This is a measure of the accuracy of the registration process. Classification accuracies will be discussed in section 6.1.3.

6.1.2.2 Classification Maps

The wall-to-wall classification for the May data set is presented as county maps in figures 6-1, 6-2, and 6-3. These maps were produced in an offset printing process from mosaics of single-class classification maps output from the Image 100. The evaluation of the classifications depicted in these maps is discussed in section 6.1.3.

6.1.3 EVALUATION RESULTS

This section presents the results of the three evaluation methods used to check the classification accuracy in TRICPS. The first evaluation, a historical data comparison, examined the area accuracy. The other two methods evaluated the classification map accuracy using test fields and randomly sampled plots.

TABLE 6-4.— COMPARISON OF IMAGE 100 CLASS AREAS BY COUNTY WITH
FOREST SERVICE FOREST LAND AND SOIL CONSERVATION SERVICE
RANGELAND STATISTICS FOR THE MAY DATA SET

Class	Image 100 Hectares (acres)	FS and SCS Hectares (acres)	% Difference of Image 100 relative to FS and SCS
MONTGOMERY			
Pine	122,151 (301,764)	105,222 (260,000)	+16.1
Hardwood	33,491 (82,754)	31,567 (78,000)	+6.1
Mixed	2,660 (6,572)	67,342 (166,400)[a]	-96.1
Range	34,087 (84,174)	39,177 (96,806)	-13.0
Other	89,796 (222,005)	39,011 (96,394)	+130.3
TOTAL AREA	282,185 (697,269)	282,319 (697,600)	-0.5
WALKER			
Pine	63,913 (157,928)	77,864 (192,400)	-17.9
Hardwood	23,435 (57,906)	23,149 (57,200)	+1.2
Mixed	612 (1,511)	29,462 (72,800)[a]	-97.9
Range	49,991 (125,135)	46,434 (114,737)	+9.1
Other	69,081 (168,190)	27,707 (68,463)	+145.7
TOTAL AREA	206,668 (510,670)	204,616 (505,600)	+1.0
SAN JACINTO			
Pine	62,401 (154,190)	80,738 (199,500)	-22.7
Hardwood	25,598 (63,253)	23,068 (57,000)	+11.0
Mixed	1,940 (4,793)	13,841 (34,200)[a]	-86.0
Range	15,522 (38,355)[b]	27,162 (67,117)	-42.9
Other	62,862 (155,330)	16,829 (41,583)	+273.5
TOTAL AREA	168,323 (415,921)	161,637 (399,400)	+4.1

[a]FS figures include abandoned cropland that is in the process of
being reforested. These areas are very sparsely populated with
trees and are not typical forest stands.

[b]Heavy rain (1.89 inches) had flooded many areas presumed to be
range. This presented difficulties in obtaining a signature to
identify range areas accurately.

TABLE 6-5.— COMPARISON OF IMAGE 100 CLASS AREAS BY COUNTY WITH
FOREST SERVICE FORESTLAND AND SOIL CONSERVATION SERVICE
RANGELAND STATISTICS FOR THE NOVEMBER DATA SET

Class	Image 100 Hectares (acres)	FS and SCS Hectares (acres)	% Difference of Image 100 relative to FS and SCS
MONTGOMERY			
Pine	114,792 (283,646)	105,222 (260,000)	+9.1
Hardwood	85,249 (210,647)	31,567 (78,000)	+170.1
Mixed[a]		67,342 (166,400)[b]	
Range	50,867 (125,691)	39,177 (96,806)	+29.8
Other	30,743 (75,966)	39,011 (96,394)	-21.2
TOTAL AREA	281,651 (695,950)	282,319 (697,600)	-0.2
WALKER			
Pine	56,040 (138,474)	77,864 (192,400)	-28.0
Hardwood	53,707 (132,707)	23,149 (57,200)	+132.0
Mixed[a]		29,462 (72,800)[b]	
Range	73,795 (182,346)	46,475 (114,737)	+58.9
Other	25,406 (62,778)	27,707 (68,463)	-8.3
TOTAL AREA	208,949 (516,305)	204,616 (505,600)	+2.1
SAN JACINTO			
Pine	68,733 (169,838)	80,738 (199,500)	-14.9
Hardwood	40,854 (100,948)	23,068 (57,000)	+77.1
Mixed[a]		13,841 (34,200)[b]	
Range	21,597 (53,366)	27,162 (67,117)	-20.5
Other	34,093 (84,242)	16,829 (41,583)	+102.6
TOTAL AREA	165,277 (408,394)	161,637 (399,400)	+2.3

[a]No signature was derived for November.

[b]FS figures include abandoned cropland that is in the process of being reforested. These areas are very sparsely populated with trees and are not typical forest stands.

TABLE 6-6.— TRI-COUNTY CLASS AREAS OF WALL-TO-WALL RESULTS

Class	May		November	
	Hectares (acres)	% Difference of Image 100 results relative to historical results	Hectares (acres)	% Difference of Image 100 results relative to historical results
Pine	248,438 (613,882)	-5.8	239,565 (591,958)	-9.2
Hardwood	82,524 (203,913)	+6.1	179,809 (444,302)	+131.2
Mixed[a]	5,211 (12,876)	-95.3	—	—
Range	100,230 (247,664)	-11.1	146,260 (361,403)	-29.7
Other	220,774 (545,525)	+164.3	90,242 (222,986)	+8.0

[a]No signature was derived for November.
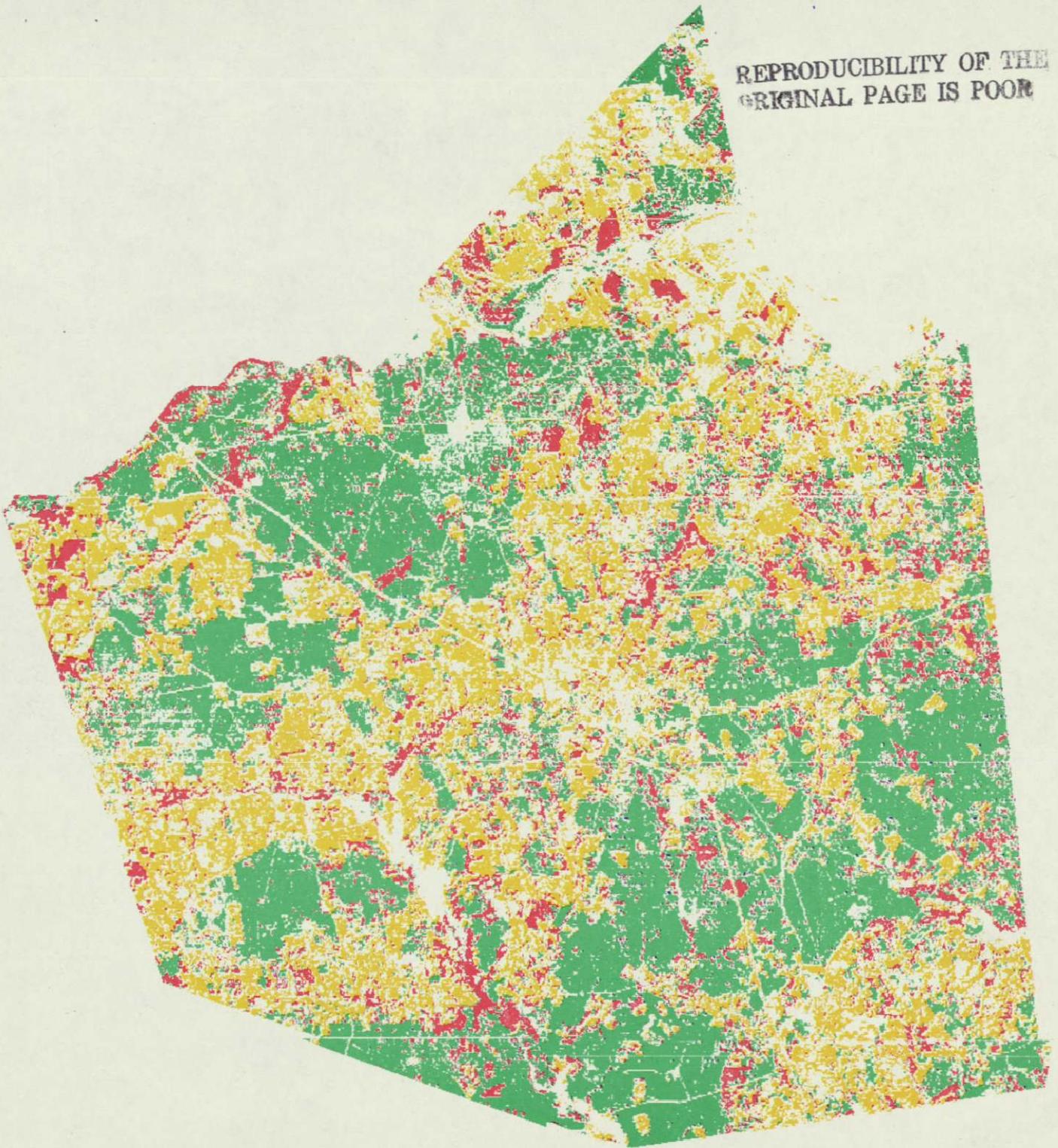
**FOREST AND RANGE INVENTORY
USING LANDSAT 1, MAY 1973
WALKER COUNTY, TEXAS**

Figure 6-1.— Classification map
of Walker County.

PINE 157,928 ACRES

HARDWOOD 57,906 ACRES

RANGE 123,526 ACRES

MIXED 1,511 ACRES

STATUTE MILES

KILOMETERS

6-10
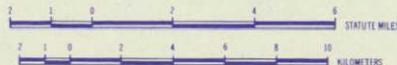
FOREST AND RANGE INVENTORY
USING LANDSAT 1, MAY 1973
MONTGOMERY COUNTY, TEXAS

PINE 301,820 ACRES

HARDWOOD 82,754 ACRES

RANGE 84,174 ACRES

MIXED 6,572 ACRES
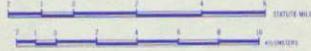
ANALYSIS BY:
  FORESTRY APPLICATIONS PROJECT
  NASA, JSC,
  HOUSTON, TEXAS 77058

PREPARED BY:
  FSO, CARTOGRAPHIC LABORATORY,
  EARTH OBSERVATIONS DIVISION,
  NASA, JSC,
  HOUSTON, TEXAS
  NOVEMBER 1975
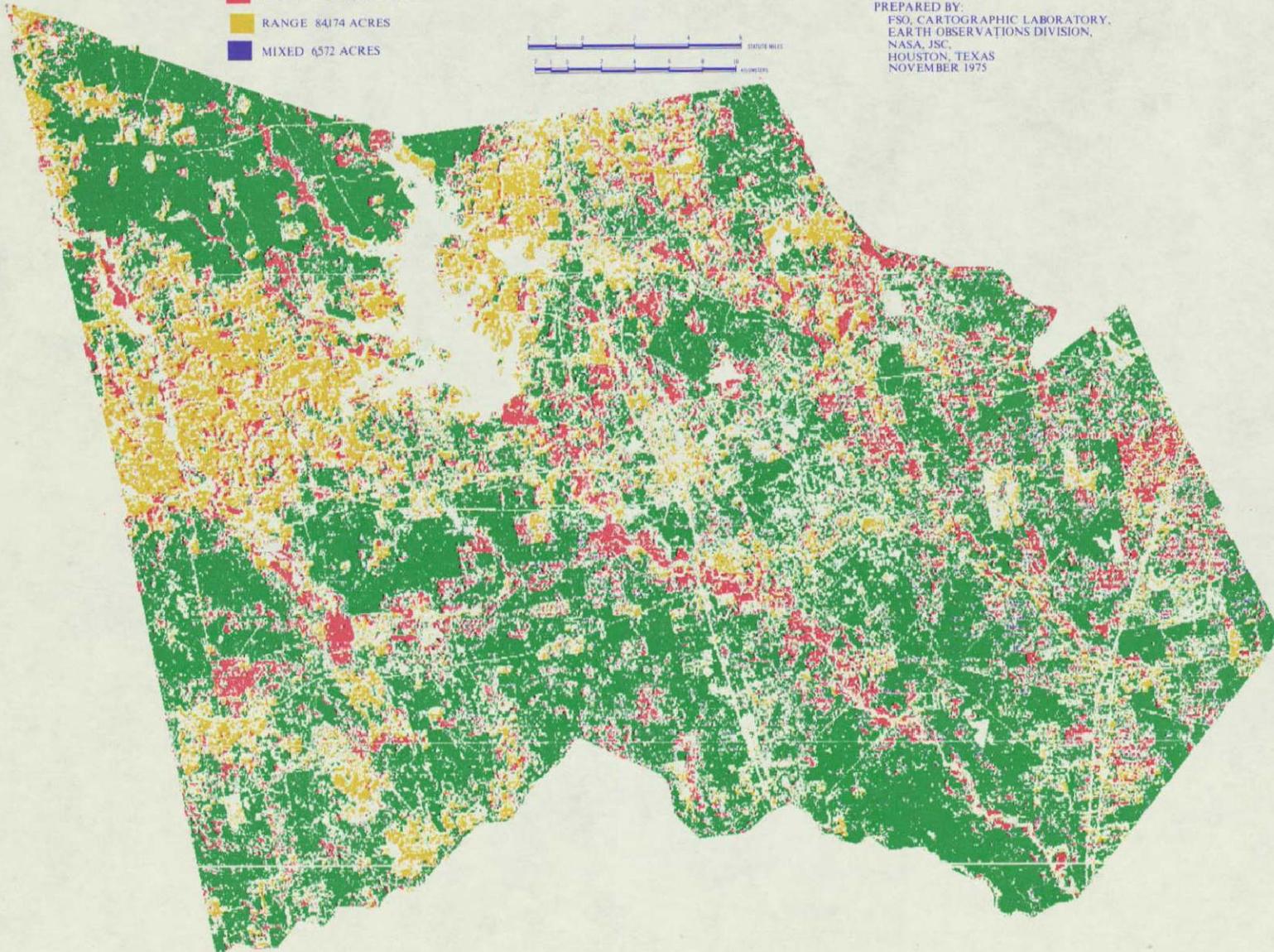
Figure 6-2.— Classification map of Montgomery County.

FOREST AND RANGE INVENTORY
USING LANDSAT 1, MAY 1973
SAN JACINTO COUNTY, TEXAS
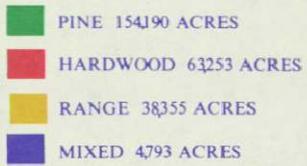
PINE 154,190 ACRES

HARDWOOD 63,253 ACRES

RANGE 38,355 ACRES

MIXED 4,793 ACRES

STATUTE MILES

KILOMETERS

ANALYSIS BY:
   FORESTRY APPLICATIONS PROJECT
   NASA, JSC,
   HOUSTON, TEXAS 77058

PREPARED BY:
   FSO, CARTOGRAPHIC LABORATORY,
   EARTH OBSERVATIONS DIVISION,
   NASA, JSC,
   HOUSTON, TEXAS,
   NOVEMBER 1975
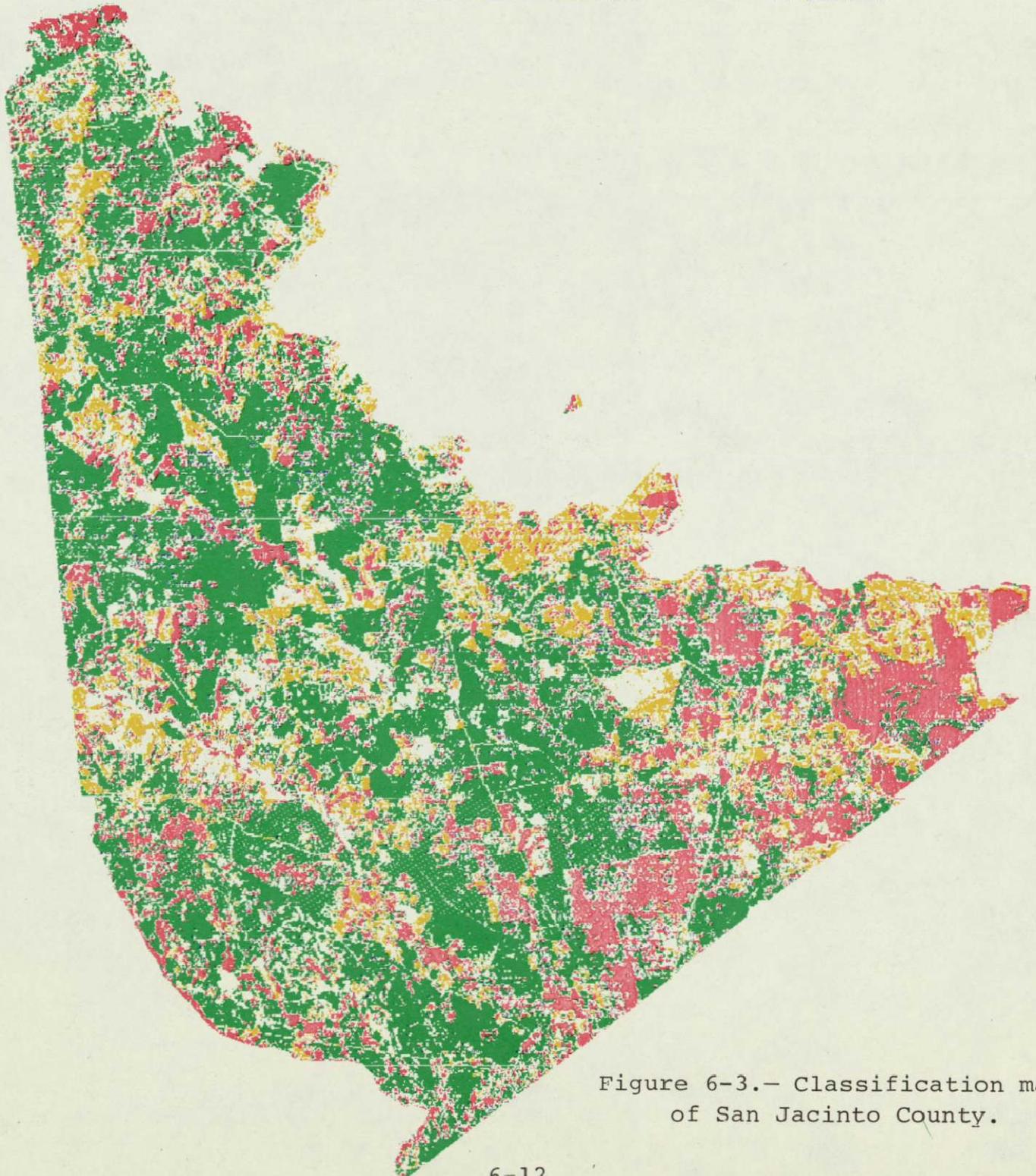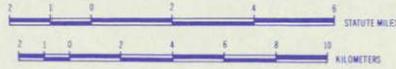
Figure 6-3.— Classification map
of San Jacinto County.

6.1.3.1 Historical Data Comparison

The areal estimates produced by classification of each data set were compared with the FS and SCS area figures* by county and by the entire TRICPS study area. The results of these comparisons for each class are shown in tables 6-4, 6-5, and 6-6 as percentage differences of Image 100 classifications relative to the FS or SCS areal estimates. For the May tri-county data, pine, hardwood, and range varied 5.1, 6.1, and 11.1 percent, respectively, from the historical data.

Some general observations can be made from these comparisons. First, mixed land in the May data set was grossly underestimated for all counties. It appeared that most of the FS-defined mixed areas were classified as "other" in TRICPS. This was not surprising to the authors because of the difficulty experienced in obtaining a mixed signature. In addition, FS figures for mixed pine/hardwood include abandoned cropland that is being reforested. These areas, which are very sparsely populated with trees, were not used as training fields for mixed pine/hardwood, and this partially accounts for the underestimation of mixed areas.

Hardwood was greatly overestimated in the November data. Because there was no mixed signature extracted for November, much of the hardwood recognized in TRICPS was really mixed.

On the average for all classes in both data sets, the greatest classification errors occurred in San Jacinto County. One explanation can be offered to account for this result in the May data set. Heavy rain flooded some range areas prior to the May Landsat overpass. This made it difficult to obtain a range signature that would recognize all range areas.

---

*The FS data (1975) used for comparison was compiled using a systematic sampling method based on aerial photography and ground measurements for forest/nonforest classification. The sampling error by county was estimated as Walker, 3 percent; Montgomery, 2 percent; and San Jacinto, 2 percent (ref. 9).

6.1.3.2 Test Field Evaluation

Test fields were selected to equal 1 percent of the area in Walker County. The test fields were chosen to be from 12 to 16 hectares (35 to 50 pixels) in size. A total of 76 fields was selected including 36 of pine, 23 of hardwood, and 17 of range. Mixed land was not evaluated because areas could not be located that were large enough to contain test fields.

The results of the evaluation are shown in the truth table (6-7). Based on a total of 3659 test pixels, the classification accuracies for pine, hardwood, and range are 95, 76, and 71 percent, respectively.

This procedure was both simple and reliable to execute. Test fields were selected only from areas positively identified from photography and large enough so there was no question that the test field was contained within the area. There are, however, some drawbacks in using this procedure. First, it does not really measure the classification map accuracy, particularly the locational accuracy. Second, problems due to transition pixels are necessarily avoided because of the discrete choice of test fields. Third, it has been shown theoretically (ref. 18, chapter 5, section 4) that this procedure tends to produce overestimates of the true classification accuracy. Fourth, there is no measure of confidence, i.e., repeatability, of the results of this evaluation. Fifth, the proportions of the features in the entire map are not being considered, a procedure which is essential in an overall map evaluation (ref. 16). Therefore, a random-sample evaluation was performed as described below.

6.1.3.3 Random-Sample Evaluation

In this procedure, one hundred 33-hectare plots throughout the study site were selected randomly on computer classification maps. Each plot was divided into nine equal size subplots, making a total

TABLE 6-7.— TRUTH TABLE SHOWING PERCENT CORRECTLY CLASSIFIED
FOR THE TEST FIELD EVALUATION OF THE MAY
DATA SET OF WALKER COUNTY

COMPUTER CLASSIFICATION

GROUND TRUTH CLASSIFICATION

	Pine	Hardwood	Range	Other
Pine 1738	1657 (95.34%)	39 (2.24%)	0 (0.00%)	42 (2.42%)
Hardwood 1101	79 (7.18%)	842 (76.48%)	7 (0.64%)	173 (15.71%)
Range 820	0 (0.00%)	8 (0.98%)	584 (71.22%)	228 (27.80%)

of 900 subplots for each date. The computer classification of each subplot was determined based on the predominant class in the subplot. Then, the plots were located on the photography using recognizable features (e.g., roads) on both the classification maps and photography. Each of the subplots was interpreted to be the class which comprised the majority of that subplot based on the photography.

Figures 6-4 and 6-5 depict the results of evaluation of 900 subplots each for the May and November data sets. The two-class probabilities* of correct classification (PCC's) are given for each class, and the overall (M-class) PCC and a corresponding confidence interval are presented for each data set (ref. 16).

The drastic difference in classification accuracy between this procedure (less than 60 percent overall for both dates) and the previous two (average about 80 percent overall) leads the authors to question the reliability of the execution. After thorough analysis, one significant source of error was found to be in registering the computer classification map to the aerial photography. An area as small as several pixels was difficult to locate with confidence. Many of the randomly chosen sample plots included in this procedure fell in boundary areas where slight registration errors would and did drastically lower the accuracy estimates. These slight registration errors were avoided in the test field evaluation method because the test fields were located well within known areas.

Another problem encountered in the execution of this procedure arose from the use of the "majority rule" to determine the identity of sample plots. Even though the majority rule is appropriate for two-class maps, it presents problems in multiclass maps, as in

---

*A multiclass map can be considered one class at a time; e.g., pine; hence, a pine/nonpine classification map — a two-class classification map. The relationship between the two-class PCC and the M-class overall PCC is shown in reference 16, in which the reasons and advantages of using this formulation are also explained.

OVERALL RESULTS OF MAY EVALUATION

	PINE	HARD WOOD	MIXED	RANGE	OTHER
PINE	225	10	0	15	10
HARDWOOD	135	127	3	27	23
MIXED	32	4	1	2	1
RANGE	28	19	0	106	39
OTHER	18	8	0	45	30

☐ CORRECT COMPUTER CLASSIFICATION
☐ INCORRECT COMPUTER CLASSIFICATION

TWO CLASS PROBABILITIES OF CORRECT CLASSIFICATION FOR MAY

	PINE	N/PINE
PINE	225	35
N/PINE	214	435

$$\frac{225 + 435}{909} = 72.6\%$$

	HARD WOOD	N/HARD WOOD
HARDWOOD	127	189
N/HARDWOOD	41	552

$$\frac{127 + 552}{909} = 74.7\%$$

	MIXED	N/MIXED
MIXED	1	39
N/MIXED	3	866

$$\frac{1 + 866}{909} = 95.4\%$$

	RANGE	N/RANGE
RANGE	106	86
N/RANGE	89	628

$$\frac{106 + 628}{909} = 80.7\%$$

	OTHER	N/OTHER
OTHER	30	71
N/OTHER	73	735

$$\frac{30 + 735}{909} = 84.2\%$$

OVERALL M-CLASS PROBABILITY OF CORRECT CLASSIFICATION FOR MAY

$$PCC = 1/2 \ [(.726 + .747 + .954 + .807 + .842) - 3]$$

$$PCC = \boxed{.538}$$

CONFIDENCE INTERVAL (CI) FOR MAY

$$CI = \left(.538 - 1.64 \sqrt{\frac{.538 \ (1 - .538)}{900}} , \ .538 + 1.64 \sqrt{\frac{.538 \ (1 - .538)}{900}}\right)$$

$$CI = \boxed{(51; 56)} \ \text{AT 0.9 CONFIDENCE LEVEL}$$

Figure 6-4.– Random sample evaluation results for the May data set.

OVERALL RESULTS OF NOVEMBER EVALUATION

☐ CORRECT COMPUTER CLASSIFICATION

☐ INCORRECT COMPUTER CLASSIFICATION

	PINE	HARD WOOD	RANGE	OTHER
PINE	232	46	7	9
HARDWOOD	113	129	16	18
RANGE	9	16	120	28
OTHER	5	37	47	17

TWO CLASS PROBABILITIES OF CORRECT CLASSIFICATION FOR NOVEMBER

	PINE	N/PINE
PINE	232	62
N/PINE	127	428

$$\frac{232 + 428}{849} = 77.7\%$$

	HARD WOOD	N/HARD WOOD
HARDWOOD	129	147
	99	474

$$\frac{129 + 474}{849} = 71.0\%$$

	RANGE	N/RANGE
RANGE	120	53
N/RANGE	70	606

$$\frac{120 + 606}{849} = 85.5\%$$

	OTHER	N/OTHER
OTHER	17	89
N/OTHER	55	688

$$\frac{17 + 688}{849} = 83.0\%$$

OVERALL M-CLASS PROBABILITY OF CORRECT CLASSIFICATION FOR NOVEMBER

$$PCC = 1/2 \ [(.777 + .71 + 855 + .83) -2]$$

$$PCC = \boxed{.586}$$

CONFIDENCE INTERVAL (CI) FOR NOVEMBER

$$MAY \ CI = \left(.586 - 1.64 \sqrt{\frac{.586 \ (1 - .586)}{849}} \ , \ .586 + 1.64 \sqrt{\frac{.586 \ (1 - .586)}{849}} \right)$$

$$CI = \boxed{(56, \ 61)} \ AT \ 0.9 \ CONFIDENCE \ LEVEL$$

Figure 6-5.— Random sample evaluation results for the November data set.

TRICPS. The majority class in a TRICPS sample plot depended upon how many classes it contained. For example, if a plot contained only two classes, the class with greater than 50-percent area was the majority; however, if a plot contained five classes, the majority class need only be greater than 20 percent. Such inconsistency allows the majority class to vary greatly in areal percentage for the computer classification versus the ground-truth data.

6.2 SAMPLING APPROACH

Based on the estimates of between-PSU variance for each class, produced from 20 random samples, the following numbers of samples are needed for each class to meet the 0.9 confidence level and ± 5 percent allowable error required in the approach (section 5.4): pine, 78; hardwood, 10; range, 39; mixed, 1; and other, 27. The largest of these, 78, was the number of samples used to estimate the wall-to-wall acreage. Fifty-eight new samples were chosen and their results added to the 20 previously selected samples to produce portions of each feature within the sample population.

These proportions were applied to the total study site to estimate areas for each feature. The results of the estimates are presented in table 6-8. Both the sampling and the wall-to-wall figures were derived from the classification of May data using the May signatures found in table 6-3.

Relative to the wall-to-wall figures, the sampling estimates differed 1.0, 0.5, 11.7, 22.1, and 8.4 percent, respectively, for pine, hardwood, mixed pine/hardwood, range, and other. The large deviation of the range estimate was unexpected because the number of samples used permitted a 5-percent allowable error with 90-percent confidence.

TABLE 6-8.— COMPARISON OF AREAS OBTAINED FROM 78 SAMPLE UNITS AND
WALL-TO-WALL AREAS OF THE MAY DATA SET

Class	Sampling hectares (acres)	Wall-to-wall hectares (acres)	% Difference of sampling relative to wall-to-wall results
Pine	245,850 (607,486)	248,438 (613,882)	−1.0
Hardwood	82,147 (202,982)	82,524 (203,913)	−0.5
Mixed	4,600 (11,367)	5,211 (12,876)	−11.7
Range	122,366 (302,362)	100,230 (247,664)	+22.1
Other	202,213 (499,662)	220,774 (545,525)	−8.4

6.3 DISCUSSION

This section presents a discussion of some possible improvements
in the procedures which were identified while carrying out TRICPS.
Specifically, modifications to the procedure were determined for
the following steps: registration and boundary input, signature
acquisition, classification, and evaluation.

6.3.1 REGISTRATION AND BOUNDARY INPUT

6.3.1.1 Size of the Segments

The breakdown of the tri-county area into segments was necessary
because of the constraints of the Image 100 memory and display.
The display was believed to accommodate 507 lines and pixels at
the outset of TRICPS. For this reason, the segments were made
500 lines and 500 pixels in size. After the segments were made,
it was found that the screen could display only 493 lines. This
presented problems in inputting boundaries and deleting areas not
within counties.

In future studies, therefore, the segment size should be reduced
to less than 490 lines and pixels. Furthermore, the segment
should be centered within the screen so there is a boundary on
all sides of the segment that contains no data. This would make
the manipulation of the boundaries easier and less time consum-
ing and would permit more accurate placement of the boundary.

6.3.1.2 Location of Control Points

To improve the registration process, control points need to be
located more accurately on the Image 100 screen. The cursor used
to identify line and pixel numbers of control points was found
after TRICPS analysis to be actually four pixels in size rather
than just one. This did not permit accurate identification of
the control points in the TRICPS registration. By reading in only

6-21

250 lines and 250 pixels, rather than 500 by 500, the cursor would really be one pixel in size; this would permit locations of control points with accuracy within one pixel instead of two. This procedure is therefore recommended for future locations of control points.

6.3.1.3 Deletion of Landsat Data

Further investigation into the inputting of maps via the input scanner determined that no one channel of Landsat data need be deleted to accommodate the map. TRICPS-manipulated data sets only contain three video channels. The new way to avoid the missing channel is as follows: Initially, the map must be input to a data channel (1, 2, 3, or 4) but then can be moved to a theme for storage (channel 5). After this is done, the data channel that was erased to accommodate the map can be restored for use in processing. Because the Image 100 is designed to handle four channels, it is worthwhile to make use of all the Landsat data, provided the data quality is acceptable. In TRICPS, however, channel 1 of the May data set contained noise which would have limited its usefulness.

6.3.2 SIGNATURE ACQUISITION

6.3.2.1 Radiometric Resolution Cell Size

Further examination and research into the proper resolution cell size indicated that more study should be done for each individual data set. The cell size should be adjusted so that the smallest standard deviation of all the classes of interest in each channel should range from 1.5 to 2.5 units (ref. 19); note that this is not the same unit of measurement on the original Landsat data but rather a transformed unit (ref. 15). This cell size would provide the optimum separability for the classes. Determining this cell resolution would require selecting training fields and extracting signatures for all classes and examining the histograms. Different cell resolutions should be tried iteratively until the proper resolution is achieved.

6.3.2.2 Truncation of Signatures

The arbitrary truncation of 1 percent of the pixels from each
end of the histograms for each channel was used to eliminate
pixels not belonging to the training class. This procedure did
provide some safeguard against stray pixels. However, a more
appropriate procedure might be actually to check the identity
of outlying pixels in the histograms for each signature. This
could be accomplished by displaying the pixels that account for
a given percentage of each tail of the histograms. Then the
identity of these pixels could be ascertained by checking against
ground truth. If they were identified as belonging to the class
of interest, they should be retained in the training signature;
otherwise, they should be deleted from the training signature.

6.3.3 CLASSIFICATION

6.3.3.1 One-Dimensional Versus N-Dimensional Classification

The Image 100 extracts two kinds of signatures, one-dimensional
and N-dimensional. The one-dimensional signature consists of the
ranges of gray levels for each channel considered independently,
i.e., a simple hypercube. This signature, which was used for
classification in TRICPS, can be depicted as a rectangle for a
two-channel case (fig. 6-6).

A more precise way to identify classes is to use N-dimensional
signatures. Figure 6-7 demonstrates how one-dimensional signa-
tures could easily overlap in the presence of two classes, A and
B. This overlap would produce pixels classified as both class A
and class B.

The N-dimensional signatures are depicted in figure 6-8 for classes
A and B. These signatures account for the interaction of data
values between channels. In the present example, the N-dimensional

One-dimensional signature
for class A

CHANNEL Y
DATA VALUES

Class A

CHANNEL X
DATA VALUES
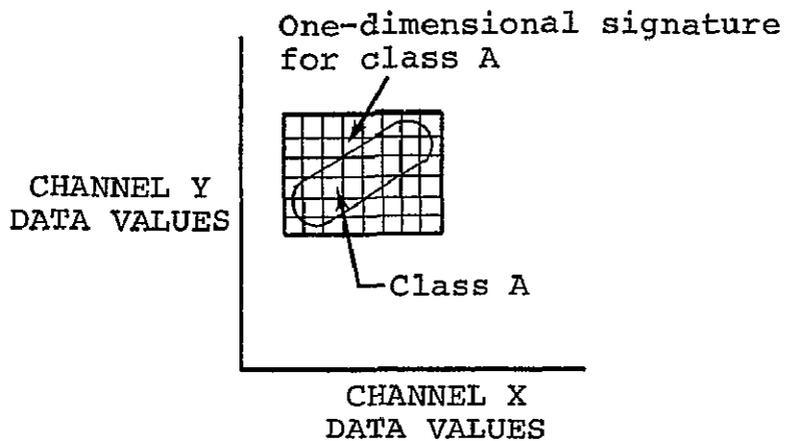
Figure 6-6.— Diagram showing one-dimensional signature.

Class A signature

Overlap

CHANNEL Y
DATA VALUES

Class B
signature

CHANNEL X
DATA VALUES
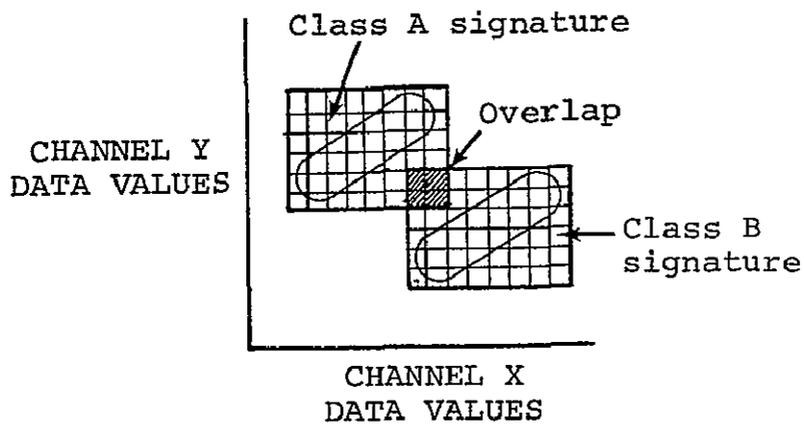
Figure 6-7.— Diagram showing overlapping of one-dimensional
signatures.
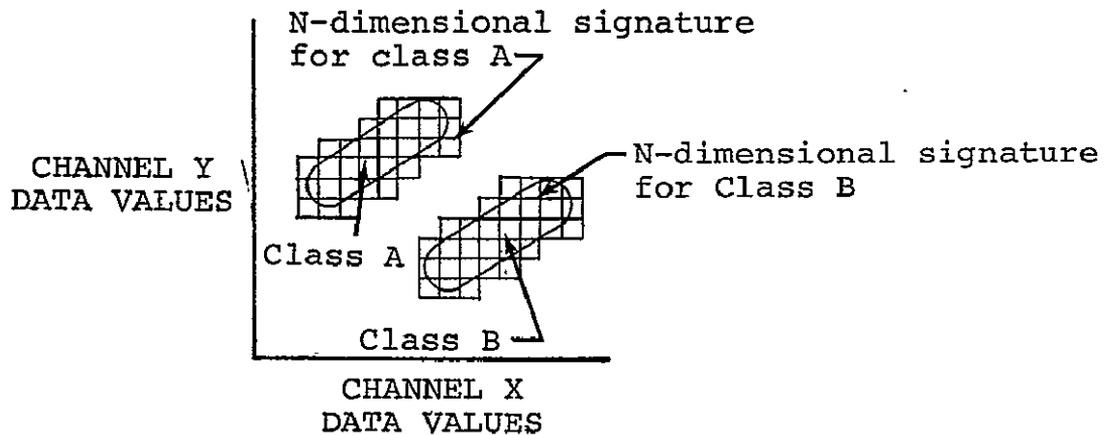
Figure 6-8.— Diagram showing N-dimensional signatures.

signatures for these classes produce no overlap between classes A and B. (Reference 15 presents a further explanation of one-dimensional versus N-dimensional signatures.)

The use of N-dimensional signatures is suggested because they allow more precise description of each class.

6.3.3.2 Resolving Overlapping Signatures

The one-dimensional signatures used in TRICPS had overlapping spectral bounds in many cases. This overlap had to be resolved in one channel for each signature pair to prevent pixels being classified as more than one class. Resolving the overlap required the truncation of the signatures that were arbitrary in TRICPS. (The need for some of this truncation might have been eliminated by using N-dimensional signatures as discussed in the previous section.)

The overlap no longer needs to be eliminated by truncation because of the addition of a nonparametric, maximum-likelihood classifier to the Image 100. This program takes any pixels that have been classified as more than one class and assigns them to one class

based on a maximum-likelihood analysis of the signatures. This procedure is recommended over the truncation of signatures.

6.3.4 EVALUATION

The difficulties encountered in the random sample evaluation led to the proposal of a new *ad hoc* evaluation procedure (ref. 20). This procedure would attempt to relieve the problems in registration of computer classification to ground truth and to provide another method of evaluating sample plots other than by majority rule.

The registration inaccuracies would be minimized in the *ad hoc* procedure by making classification maps at a larger scale than in TRICPS and including all classes in one map.* Both of these improvements would make registration a more reliable process. Also, the sampling procedure would be carried out by a hierarchy of primary and secondary sampling units. The primary units are large enough so that landmarks can be identified for registration; then, the secondary sample units are located in the primary units; and finally, the nine nearest neighbors (ref. 20) for each potential secondary sample unit would be examined to find the one with the smallest error. This procedure would account for misregistration of one pixel in any direction. Extension to more than one-pixel misregistration is possible.

The inaccuracies arising from comparison by majority rule would be reduced by using a new error measure based on the difference of proportions of the classes. This method is more discriminatory than the majority rule. Reference 20 presents a discussion of this new measure.

---

*The maps used in TRICPS contained only one class each.

7. RESOURCE REQUIREMENTS

7.1 MANPOWER

The TRICPS was performed by three project personnel, all having photointerpretation and automatic data processing experience. One year of work was performed on data preparation and analysis prior to final reporting. In addition, one man-year-equivalent of support services, including cartographic and photographic support, was needed.

7.2 EQUIPMENT

Equipment requirements throughout the 1-year data analysis period were approximately as follows:

● Image 100 - 12 hours per week

● PMIS-DAS - 54 hours

● LARSYS terminal - 2 hours

● Kargl reflecting projector - 60 hours

● Zoom transferscope - 16 hours

7.2.1 IMAGE 100

The Image 100 is an interactive computer system. This system features (1) a PDP-11/45 computer, (2) image analyzer console, (3) line printer, (4) graphic display terminal, (5) magnetic tape drives, (6) input scanner unit, and (7) solid-state refresh memory (fig. 7-1).

At present, four channels of MSS data in Landsat, LARSYS, or Universal format can be input. The console screen displays 493 lines by 507 pixels (ref. 15). The input scanner unit can be used to digitize imagery or maps. Data such as administrative boundaries on a map can then be input to the console screen as the base to which the Landsat data can be registered.
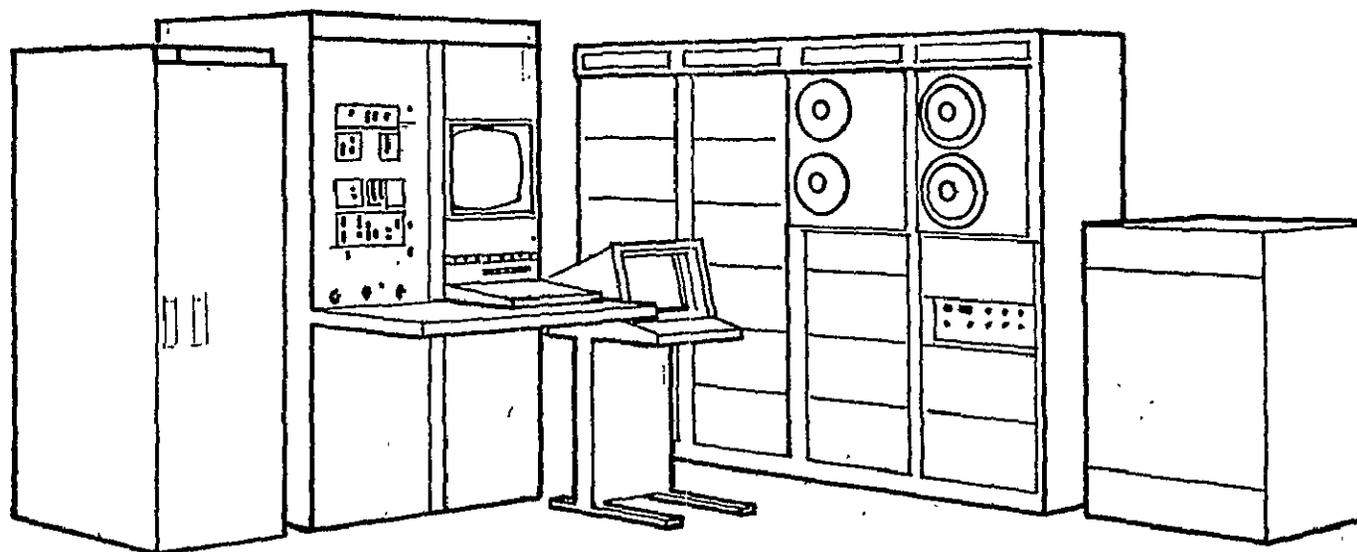
Figure 7-1.- Image 100.

One class at a time is considered in classification. Upper and lower spectral limits for each channel can be established, and all pixels falling within these limits are then classified into that class. Classification results may be stored and output to magnetic tape or on the line printer. For more information see reference 15.

7.2.2 PMIS DAS

The PMIS DAS is a data processing tool used in the screening and editing of digital imagery collected by various airborne and orbital multispectral scanners. The system features (1) a SEL 810B computer, (2) tape control units, (3) CDC tape drives, (4) line printer, (5) card reader, (6) card punch, (7) control console, (8) data input monitor, (9) ASR 33 teletype, (10) Singer film recorder, and (11) mass refresh memory.

At present, multichannel data in Landsat, LARSYS, or Universal format can be input. Three channels at a time can be displayed and enhanced on the control console. Data can be reformatted to Landsat, LARSYS, or Universal format as desired. Output can be in the form of magnetic tape or film transparencies. For more information see reference 21.

7.2.3 LARSYS TERMINAL

The LARSYS terminal permits access to LARS computer programs, which are used in the registration of Landsat data.

7.2.4 KARGL REFLECTING PROJECTOR

The Kargl reflecting projector makes it possible to register photographic transparencies and theme classification maps which have different scales.

7.2.5 ZOOM TRANSFERSCOPE

The Zoom transferscope performs essentially the same operation as the above — a registration of photographic transparencies and theme classification maps that have different scales.

7.3 PROCESSING TIME UTILIZATION

Preprocessing for this study required approximately 44 man-hours. Registration preparation and registration per segment (500-line by 500-pixel area) took 2 hours of Image 100 time and 0.1 hour on the LARSYS terminal.

Training field selection and verification required 105 hours. Each segment took from 4 to 6 hours depending on the complexity. of the classes.

Signature acquisition per county for the four classes required 24 hours. Actual classification took 5 hours (0.24 hour per segment). Outputting the results to the line printer and to magnetic tapes took an additional 5 hours.

The sampling procedure required 40 hours Image 100 time, 3 hours ; of sample unit selection, and 5 hours to compile the results.

The test field evaluation for the Walker County May data set required a total of approximately 25 hours. Photointerpretation and registration using the Zoom transferscope took 16 hours, and pixel counts and tabulation of the results took 9 hours.

Random-sample evaluation of classification results took a total of 80 hours for the two data sets. To locate and evaluate the plots on the classification maps required 17 hours; to locate the plots on the photography required 17 hours. Photointerpretation

of the plots took 30 hours. The comparison of computer classifi-
cation and photointerpretation required 6 hours. The application
of PCC and confidence interval to the results required 10 hours.

For a more detailed breakdown of time utilization, refer to the
appendixes.

8. CONCLUSIONS AND RECOMMENDATIONS

8.1 CONCLUSIONS

The following specific objectives of the study were accomplished:
(1) a computer-aided classification procedure to inventory forest-
land and rangeland by administrative boundaries was developed and
tested, and (2) a sampling procedure to inventory large areas
using a subset of the data was developed and tested.

In response to the above specific objectives, the main purpose of
the study was to design procedures, test their feasibility, and
work out possible solutions to problems that developed. Specific
procedures were developed for (1) registering Landsat data to
topographic maps, (2) inputting boundaries from a topographic map
onto the Landsat data, (3) obtaining class signatures from train-
ing fields, (4) developing computer classification procedures for
both the wall-to-wall approach and the sampling approach, (5)
developing a method for producing county classification maps from
computer classification printouts, and (6) designing techniques
for evaluation of the classification results. (Refer to section 5
and appendixes B through G.)

Although the size of the tri-county study site did not warrant use
of a sampling approach, the procedure was found to be readily
usable and would be useful in preparing for future state, regional,
or national inventories where a sampling approach would, for eco-
nomic reasons, be unavoidable.

The mapping of mixed pine/hardwood areas was concluded to be
infeasible using the procedures of TRICPS. The location of enough
homogeneous mixed areas sufficiently large to produce a meaningful
signature was not possible. Most areas identified as mixed by the
FS or through photointerpretation contained small stands of pure
pine or hardwood, resulting in a signature characteristic of both
pine and hardwood.

For pine, hardwood, and range, the training fields selected in each county were very similar in spectral bounds. This implies that any of the county signatures could have been used to classify another county or the entire study site, resulting in similar classifications.

The May and November data sets produced similar classification accuracies except in the case of hardwood. The November data set grossly overestimated hardwood, based on a comparison with historical data. This indicates that the signature used for hardwood in the November data set was unable to distinguish hardwood from other features in the scene. It can be concluded from this result that the November data are not appropriate for classification of forest scenes for this study site.

The results, in the form of computer classification maps or classification statistics, were evaluated by three different methods. The evaluation with historical data from the FS and SCS varied considerably among pine, hardwood, mixed pine/hardwood, range, and other. The areal figures for the May data set varied 5.8, 6.1, 95.3, 11.1, and 164.3 percent, respectively, for pine, hardwood, range, mixed pine/hardwood, and other, relative to FS and SCS statistics; the November data set showed differences of 9.2, 131.2, 29.7, and 0.8 percent, respectively, for pine, hardwood, range, and other. The test field evaluation produced the highest classification accuracy figures: 95, 76, and 71 percent for pine, hardwood, and range, respectively, for the Walker County May data set. Classification accuracy figures for the random-sample evaluation were much lower than those for the test field evaluation. The overall multiclass PCC was 53.8 and 58.6 percent for the May and November data, respectively. For the May data set, the "two-class accuracies" were 72.6, 74.7, 95.4, 80.7, and 84.2 percent for pine, hardwood, mixed pine/hardwood, range, and other, respectively. The November data set produced two-class accuracies of 77.7, 71.0, 85.5, and 83.0 percent for pine, hardwood, range, and other, respectively.

8.2 RECOMMENDATIONS

The recommendations for improvements in the TRICPS procedure can be broken down into registration and boundary input, signature acquisition, classification, and evaluation. These topics are discussed in the following subsections and in section 6.3.

8.2.1 REGISTRATION AND BOUNDARY INPUT

a. Segments for processing with the Image 100 should be less than 490 lines and pixels in size (because of system constraints) and should be read in so that they are centered on the display screen. This will maximize the accuracy of boundary inputs and the determination of statistics by administrative boundaries as well as decrease the time necessary to input the boundary.

b. Control points for registration should be located with the Landsat data enlarged by two times on the Image 100 display screen. This will provide accurate location of the points because the cursor used is actually two lines by two pixels in size. Enlargement of the data makes the cursor equivalent to one pixel.

c. The input of boundaries can be accomplished without deletion of a data channel as in TRICPS. All Landsat channels should be used unless there is a specific reason why one is not acceptable.

8.2.2 SIGNATURE ACQUISITION

a. The radiometric resolution cell size for data in the Image 100 refresh memory should be adjusted after examining the signatures of the classes to be classified. The cell size should be adjusted in each channel so that the signatures have a standard deviation between 1.5 and 2.5.

b. The tails of the histograms of training field signatures should not be truncated without checking the identity of the

training pixels falling within the histogram tails. Arbitrary truncation of histograms may remove unwanted outlying pixels but may also delete pixels actually belonging to the training class.

8.2.3 CLASSIFICATION

a. The N-dimensional signatures extracted on the Image 100 should be used for classification instead of one-dimensional signatures. N-dimensional signatures more precisely describe the training pixels, which helps to reduce the overlap of signatures.

b. Signatures no longer need to be truncated arbitrarily even if they have overlapping spectral bounds. A nonparametric, maximum-likelihood classifier programmed into the Image 100 is recommended for assigning dually classified pixels to one class based on statistical properties of the overlapping signatures.

8.2.4 EVALUATION

The *ad hoc* evaluation procedure (ref. 20) is recommended for determining the accuracy of computer classification based on randomly sampled plots. This procedure minimizes the inaccuracies in registering computer classification to ground truth information. Also, this procedure replaces the use of the majority rule to determine the identity of sample plots, hence reducing the associated errors.

9. REFERENCES

1. Forestry Applications Exploratory Studies Project, a Preliminary Plan. NASA/JSC Tech. Report JSC-09420, Houston, Texas, Sept. 1974.

2. Heller, R. C., et al.: Evaluation of ERTS-1 Data for Inventory of Forest and Rangeland and Detection of Forest Stress. Pacific Southwest and Rocky Mountains Forest and Range Experimental Stations, FS, U.S. Department of Agriculture, Type III Final Report, Contract S-70251-AG, sponsored by NASA, Goddard Space Flight Center, Code 430 GSFC, Greenbelt, Maryland, Dec. 1974.

3. Erb, R. B.: A Compendium of Analysis Results of the Utility of ERTS-1 Data for Land Resources Management. NASA ERTS-1 Investigation (ER-600), NASA SP-347, JSC-08445, June 1974.

4. FS: Assessment (draft), the Nation's Renewable Resources, 1975, as Required by the Forest and Rangeland Renewable Resources Planning Act of 1974. U.S. Department of Agriculture, Washington, D.C., 1975.

5. FS: Program (draft), the Forest Service Program for the Nation's Renewable Resources, Draft Environmental Statement, Renewable Resource Program, 1977 to 2020, as Required by the Forest and Rangeland Resources Planning Act of 1974. U.S. Department of Agriculture, Washington, D.C., 1975.

6. Anderson, J. R.; Hardy, E. E.; and Roach, J. T.: A Land-Use Classification System for Use with Remote-Sensor Data. Circular 671, Washington, D.C., 1972.

7. Society of American Foresters: Forest Cover Types of North America. Washington, D.C., 1954.

8. Ford-Robertson, R. C.: Terminology of Forest Science, Technology Practice and Products. Society of American Foresters, Washington, D.C., 1971.

9. Sternitzke, H. S.: East Texas Piney Woods. Southern Forest Experiment Station, FS, U.S. Department of Agriculture, New Orleans, La., 1967.

10. ERTS Reference Manual. General Electric, Valley Forge Space Center, Philadelphia, Pa., 1974.

11. Dillman, R. D.; and Kan, E. P.: Timber Type Separability in Southeastern United States on Landsat-1 MSS Data. Lockheed Electronics Co., Inc., Tech. Report, Houston, Texas, July 1975.

12. Sternitzke, H. S.: Personal Communication of Current Texas County Forest Land Statistics. Southern Forest Experiment Station, FS, U.S. Department of Agriculture, New Orleans, La., 1975.

13. Soil Conservation Service: Texas Soil and Water Conservation Needs Inventory. U.S. Department of Agriculture, Temple, Texas, 1970.

14. Personal Communication from Huntsville, Texas, Weather Station, NOAA, concerning rainfall data for May 1973.

15. Image 100, A User Manual. General Electric Company, Daytona Beach, Florida, June 1975.

16. Kan, E. P.: Multi-Class Map Accuracy Evaluation. Lockheed Electronics Co., Inc., LEC-7936, Houston, Texas, March 1976.

17. Freese, F: Elementary Statistical Methods for Foresters. Agriculture Handbook 317, FS, U.S. Department of Agriculture, March 1974.

18. Fugonaga, K.: Introduction to Statistical Pattern Recognition. Academic Press, New York, 1972.

19. Eppler, W. G.: Personal Communication, 1975.

20. Kan, E. P.: An Ad Hoc Map Evaluation Procedure. Lockheed Electronics Co., Inc., LEC-8278, Houston, Texas, April 1976.

21. Multispectral/Microwave Data Analysis Station, Operation, and Maintenance Technical Manual. NASA/JSC Tech. Memo. JSC-13531, Houston, Texas, May 1971.

APPENDIX A

RADIOMETRIC CELL RESOLUTION STUDY

APPENDIX A

RADIOMETRIC CELL RESOLUTION STUDY

To determine the effects of resolution cell size on classification
results, classifications using cell sizes of 32, 64, 128, and 256
were examined. The resolution cell size limits the range of gray
values. For example, with a cell size of 32 all gray values would
be placed from 0 through 31.

Landsat gray value ranges are from 0 through 127 for channels 4,
5, and 6, and 0 through 63 for channel 7. The data were initially
read in refresh memory so that gray values were from 0 through 255
in all channels. The data gray values are divided by the cell
size to determine the number of data gray values assigned to each
cell, as tabulated below.

Resolution Cell Size	Number of Gray Values in Cell
32	8
64	4
128	2
256	1

A subset of the TRICPS data (200 pixels by 200 pixels) was selected,
which had extensive ground truth in the form of FS compartment maps
and aircraft photography. A detailed photointerpretation of the
corresponding aircraft photos was compiled. These areas were men-
surated using the Dell Foster digitizing system to obtain class
acreages. This provided a ground base by which to compare
acreages obtained from Image 100 classification.

Two training fields each for pine and hardwood were selected and
signatures were extracted using resolution cell sizes of 32, 64,
128, and 256. Using a cell size of 32, pine and hardwood could

not be differentiated in the May data set. With a resolution cell
size of 64, the differences between the ground base acreage and
the classification acreages were 0.4 percent for pine and 1.2 per-
cent for hardwood. Using resolution cell sizes of 128 and 256,
the accuracies were not improved. Since the size 64 cell resolu-
tion produced acreages which correlated well with the ground base,
and higher resolution did not improve the accuracy, this resolu-
tion was used in the TRICPS.

APPENDIX B

REGISTRATION PREPARATION

APPENDIX B

REGISTRATION PREPARATION

In the TRICPS, registration was the process of removing the skew
from the Landsat data and rotating the data to correspond to GS
topographic maps. This allowed county boundaries to be super-
imposed on the Landsat data.

1. The scale of the base map needed to represent the study area
   was first determined. Since the Image 100 input scanner is
   limited to an 8- by 8-inch format for the base map, scales
   vary according to size of the study area. For the TRICPS
   area of interest, a scale of 1:617,000 was determined to fit
   the 8- by 8-inch format.

2. A 1:250,000 GS topographic map of the study area was reduced
   to a color positive transparency at the determined scale of
   1:617,000.

3. Segments of Landsat-1 data to be registered were filmed at an
   approximate scale of 1:148,000 using the PMIS DAS. This scale
   was the result of the 9- by 9-inch format of the DAS trans-
   parency, which contained 250,000 pixels. (The transparencies
   are used for locating ground control points for registration.)
   ~54 hours*

4. Between 10 and 20 points (road, pipeline, and railroad inter-
   sections) were located and marked on both the topographic base
   map and the DAS transparency (1:148,000).

5. A one-tenth-inch grid overlay was registered to the base map
   transparency (1:617,000), and x and y coordinates of the points
   were recorded to two decimal places. All future overlays were
   pin registered to the base map transparency.

---

*Approximate time allotment per segment.

6. Landsat data were displayed at a normal scale (without zoom or compression) on the Image 100 screen. Using the scaled cursor on the Image 100, the coordinates of Landsat data points previously selected were determined.

7. Line and pixel values from the Landsat data, and x and y coordinates from the grid were input to the least-squares-fit computer program. Points with large errors in the least-squares-fit were eliminated on successive computer runs until root-mean-square sample error was less than two pixels (134 meters). 0.5 hour*

8. Coefficients output from the least-squares-fit computer program and starting line and pixel point were used as inputs to the Image Rotation Factor program. (Starting line and pixel position were selected so that the least number of data segments were needed to cover the study area.)** After the initial segment position was determined, the starting line and pixel for the remaining segments were determined. 1.5 hours*

---

*Approximate time allotment per segment.

**The Image 100 is limited to a maximum input for processing of 512 lines by 512 pixels. Hardware restraints for the Image 100 allow the display of 493 lines and 507 pixels; however, all pixels within the 512 by 512 limits were statistically processed. For convenience the data were divided into fourteen segments of 500 lines by 500 pixels.

APPENDIX C

REGISTRATION

APPENDIX C

REGISTRATION

1. The Image 100 camera was adjusted until the base map transparency could be input to the Image 100 screen at the proper scale. To accomplish this, the one-tenth-inch registration grid was placed on registration pins on the input scanner and viewed on the Image 100 video display. The scaled cursor was adjusted to a square 274 by 274 pixels. The camera scale was then adjusted until the cursored area represented 1 square inch on the grid.* ~0.5 hour**

2. Using the beginning and ending line and pixel and the rotation factor from the Image Rotation program, a segment of Landsat data was read into the Image 100 memory. ~0.1 hour**

3. The grid was then removed from the input scanner, and the base map transparency and the registration overlay were placed on the registration pins.

4. The base map and overlay were simultaneously moved until the corners of the first segment on the overlay corresponded with the corners of the segment displayed on the Image 100 video.

---

*It was determined through a series of calculations that if the minimum dimension of the original Landsat pixel (57.2 meters) was retained as the size of the resulting pixel, a distance of 1 inch on the 1:617,000 base transparency needed to be adjusted to fit 274 pixels on the video screen. Hence, when Landsat data were read into the Image 100 memory in the registration process, only duplication of the across-line dimension needed to be effected. In fact, 352 lines and 500 pixels/line of the original Landsat data were to be transformed into 500 lines by 500 pixels of Image 100 data. With this magnitude of pixel resolution, the printer would output maps at a scale roughly 1:115,000 (50 lines/inch and 50 pixels/ inch on the printer).

**Approximate time allotment per segment.

The overlay was removed and final adjustments were made manually with a joystick to line up roads, pipelines, etc., by alternately viewing the base map and the video image. ~1.0 hour*

5. When the alignment of the two images was satisfactory, the camera image was input to channel 1 and was in registration with the other data channels.

*Approximate time allotment per segment.

APPENDIX D

INPUTTING BOUNDARIES

APPENDIX D

INPUTTING BOUNDARIES

1. A tape containing a registered segment was restored on the
    video screen of the Image 100. The registered topographic
    map was on channel 1.

2. The Irregular Cursor program was used to input the county
    boundary from the map onto a theme. Putting in the boundary
    involved tracing the boundary on the map and placing points
    so that a smooth boundary line would result. ~0.5 hour*

3. Lines were drawn between the points by hitting "ATTENTION."
    The irregular cursor accepts up to 29 points.

4. To close the boundary, the execute button was pressed, and .
    the area cursored was automatically blocked in and stored on
    a theme.**

---

*Approximate time allotment per segment.

**If the data inside the boundary are needed, the inverse of the
boundary is loaded on a theme.

APPENDIX E

SIGNATURE ACQUISITION

APPENDIX E

SIGNATURE ACQUISITION

1. Registered Landsat segments were converted to LARSYS format on the Image 100.

2. Segments were filmed on the PMIS DAS in simulated color IR (channel 7 on red, channel 5 on green, channel 4 on blue) with biases and gains set to enhance forest and range features.

3. Segments were photointerpreted: pine, hardwood, range, and mixed pine/hardwood areas were delineated. ~1.0 hour*

4. Training fields were selected from delineated areas on the Landsat data. ~1.5 hour*

5. Fields were verified with aircraft data (MX197 at 1:120,000, MX239 at 1:60,000). ~1.5 hours*

6. Verified fields were located on the Image 100 screen with the cursor. ~0.2 hour*

7. The scaled cursor was adjusted to 6 by 6 pixels (approximately 11.7 hectares, or 29 acres); it could be smaller if necessary because of narrow range or hardwood areas. ~0.05 hour*
    The original size proposed was 5 by 5 but the cursor could not be suitably adjusted.

8. The area within the cursor was stored on a theme for future use as a training field. ~0.05 hour*

9. Field coordinates (line and pixel number) of the training fields were obtained from the scaled cursor program. ~0.1 hour*

10. Histograms of each training field were obtained using one-dimensional histogram display. ~0.1 hour*

11. The number of occurrences for each gray level was obtained using the one-dimensional histogram list. ~0.1 hour*

---

*Approximate time allotment per segment.

12. The frequency of gray levels was tabulated for all classes. ~0.1 hour*

13. Gray levels containing pixel counts with less than 1 percent frequency were truncated, producing the signatures used in classification.

14. In the event that any two signatures overlap in all channels, the overlap must be resolved in at least one channel to prevent pixels from being classified as belonging to more than one class.

*Approximate time allotment per segment.

APPENDIX F

CLASSIFICATION

1. The signature for each class was input to the Image 100 using the Cluster Synthesis Program, which loads the classification for each signature on a theme.* ~0.15 hour**

2. Areas outside the tri-county area were subtracted from the classification on each theme. ~0.05 hour**.

3. Using the Theme Area Program, theme areas of the classification results were obtained. ~0.05 hour**

4. Theme prints of the classifications were made on the line printer. ~0.1 hour**

5. Themes were written to tape via the Theme I/O Program. . ~0.1 hour**

---

*Each theme on the Image 100 was represented by one bit of an eight-bit word. One classification could be stored on each theme. Eight themes were available on the Image 100. Reference 15 provides additional program information.

**Approximate time allotment per segment.

APPENDIX G

EVALUATION OF CLASSIFICATION RESULTS

APPENDIX G

EVALUATION OF CLASSIFICATION RESULTS

G.1 TEST FIELD EVALUATION

1. Test fields (35 by 50 pixels in size), comprising a 1-percent sample, were chosen in known areas of pine, hardwood, and range on color infrared (IR) photography (1:120,000). ~6 hours*

2. Using the Zoom transferscope, color IR photography with the located test fields was registered to the computer classification maps, and the test fields were marked on them. ~10 hours*

3. Test fields were interpreted on the classification maps by counting the number of pixels in each class. ~7 hours*

4. The classification map interpretation was compared to the photointerpretation of the photography, and the number of correctly classified subplots was tabulated. ~2 hours*

G.2 RANDOM-SAMPLE EVALUATION

1. Using a 10- by 10-pixel plot, 100 plots were randomly located on the computer classification map of each class. ~6 hours*

2. Each plot was arbitrarily divided into nine subplots.

3. The computer maps were examined to determine which class comprised the majority of the subplot, and the subplot was assigned to that class. ~4 hours*

4. The computer maps were registered to photography using the Kargl reflecting projector, which enlarges, reduces, or rotates an image. ~8 hours*

---

*Approximate time allotment.

5. The plots were located on the photography. ~2 hours*

6. Each subplot was photointerpreted and assigned to the dominant class. ~16 hours*

7. The computer classification was compared to the photointerpretation, and the number of correctly classified subplots was tabulated. ~4 hours*

8. Probability of correct classification and confidence interval factors were applied to the results to check reliability. ~8 hours*

*Approximate time allotment.